# REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188



| 1. REPORT DATE (DD-MM-YYYY) | 2. REPORT TYPE | 3. DATES COVERED (From - To) |
|---|---|---|
| 27-08-2015 | Final Report | 18-Jan-2012 - 17-Jan-2015 |

4. TITLE AND SUBTITLE
Final Report for "DESPIC: Detecting Early Signatures of Persuasion in Information Cascades"

5a. CONTRACT NUMBER
W911NF-12-1-0037

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHORS
Alessandro Flammini, Filippo Menczer, Qiaozhu Mei, Sergey Malinchik

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Indiana University at Bloomington
509 E 3RD ST

Bloomington, IN 47401 -3654

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
61766-NS-DRP.59

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
The goal of DESPIC project was developing a technological infrastructure to automatically detect orchestrated campaigns on social media in their early stage of diffusion. Such campaigns include rumors, spread of misinformation, persuasion attempts, and advertising.
We designed and implemented a distributed infrastructure for the efficient collection, archival and retrieval of Twitter data, and a framework for clustering messages in topically coherent memes in a streaming scenario.
Our infrastructure has served as the basis for the development of machine learning infrastructures to discriminate

15. SUBJECT TERMS
infrastructure for Twitter data collection, archival and retrieval. clustering of streaming social media data. Detection of Social Bots. Detection of promoted and grass root trending topics. Detection of Rumors. Prediction of burstiness and popularity of hashtags

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 15. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | Alessandro Flammini |
| UU | UU | UU | UU | | 19b. TELEPHONE NUMBER 812-856-1830 |

## Report Title

Final Report for "DESPIC: Detecting Early Signatures of Persuasion in Information Cascades"

## ABSTRACT

The goal of DESPIC project was developing a technological infrastructure to automatically detect orchestrated campaigns on social media in their early stage of diffusion. Such campaigns include rumors, spread of misinformation, persuasion attempts, and advertising.
We designed and implemented a distributed infrastructure for the efficient collection, archival and retrieval of Twitter data, and a framework for clustering messages in topically coherent memes in a streaming scenario.
Our infrastructure has served as the basis for the development of machine learning infrastructures to discriminate between naturally trending and promoted content, the identification of social bots, the identification of rumors, and the prediction of burstiness and popularity of memes.

---

## Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

### (a) Papers published in peer-reviewed journals (N/A for none)

| Received | Paper |
|---|---|
| 08/19/2015 57.00 | Giovanni Luca Ciampaglia , Prashant Shiralkar , Luis M. Rocha , Johan Bollen , Filippo Menczer , Alessandro Flammini. Computational Fact Checking from Knowledge Networks, PLoS ONE, (06 2015): 128193. doi: |
| **TOTAL:** | **1** |

**Number of Papers published in peer-reviewed journals:**

---

### (b) Papers published in non-peer-reviewed journals (N/A for none)

| Received | Paper |
|---|---|
| 08/20/2014 52.00 | Azadeh Nematzadeh, Emilio Ferrara , Alessandro Flammini, Yong-Yeol Ahn. Optimal Network Modularity for Information Diffusion, Physical Review Letters, (08 2014): 88701. doi: |
| 08/23/2013 26.00 | MD Conover, C Davis, E Ferrara, K McKelvey, F Menczer , A Flammini. The geospatial characteristics of a social movement communication network, PLoS ONE, (03 2013): 0. doi: |
| 08/23/2013 27.00 | MD Conover, E Ferrara, F Menczer, A Flammini. The Digital Evolution of Occupy Wall Street, PLoS ONE, (05 2013): 0. doi: |
| **TOTAL:** | **3** |

**Number of Papers published in non peer-reviewed journals:**

---

## (c) Presentations

IU presentations:

Emilio Ferrara:

"Predicting human behaviors in techno-social systems: fighting abuse and illicit activities"

University of Southern California [invited talk]. Los Angeles, CA (US). April 2015

Syracuse University [invited talk]. Syracuse, NY (US). April 2015.

Northeastern University [invited talk]. Boston, MT (US). March 2015.

New Jersey Institute of Technology. Newark, NJ (US). March 2015.
5. DARPA ADAMS/SMISC Meeting [invited talk]. Arlington, VA (US), Mar. 2015.

"The rise of social bots: fighting deception and misinformation on social media".

Indiana University [invited talk]: Network Science talks. Bloomington, IN (US). Jan. 2015.

Texas A&M University [invited talk]. Austin, TX (US). Sept. 2014.

Denmark Technical University (DTU) [invited talk]. Copenhagen, Denmark. Sept. 2014.

Giovanni Luca Ciampaglia:

“Computational fact checking from knowledge networks”

International Conference on Network Science (NetSci'15), Zaragoza, Spain. June 4th 2015. Oral presentation.

Leibniz Institute for the Social Sciences, Cologne, Germany. March 5th 2015. Invited talk.

CRASSH Symposium on Conspiracy Theories and Democracy. University of Cambridge, Cambridge, UK, March 3rd 2015. Invited talk.

Network Science Colloquium, Indiana University, Bloomington, IN, October 13th 2014. Oral presentation.

Filippo Menczer:

“The spread of misinformation in social media." Keynote, AAAI Spring 2015 Symposium on Sociotechnical Behavior Mining, Stanford University, March 2015

University of Michigan:

Qiaozhu Mei:

"The Foreseer: Data Mining Of the People, By the People, and For the People," LinkedIn, Mountain View, CA, March 19th, 2015.

"RumorLens: Early Detection and Analysis of Rumors in Social Media," Facebook, Menlo Park, CA, March 19th, 2015.

"A New Age of Text Mining: Recent Developments and Future Directions," Bloomberg, New York, NY, August 22nd, 2014.

**Number of Presentations:** 0.00

## Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

| Received | Paper |
|---|---|

**TOTAL:**

**Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):**

## Peer-Reviewed Conference Proceeding publications (other than abstracts):

| Received | Paper |
|---|---|
| 08/19/2014 38.00 | Onur Varol, Filippo Menczer. Connecting Dream Networks Across Cultures, Proceedings of the companion publication of the 23rd international conference on World wide web, 2014. 07-APR-14, . : , |
| 08/19/2014 41.00 | Onur Varol, , Emilio Ferrara, Christine Ogan , Filippo Menczer, Alessandro Flammini. Evolution of Online User Behavior During a Social Upheaval, ACM Web Science 2014. 23-JUN-14, . : , |
| 08/19/2014 44.00 | Judy Qiu, Xiaoming Gao . Social Media Data Analysis with IndexedHBase and Iterative MapReduce, Proceedings of the 6th Workshop on Many-Task Computing on Clouds, Grids, and Supercomputers (MTAGS 2013) at Super Computing 2013. Denver, CO, USA, November 17th, 2013. 17-NOV-13, . : , |
| 08/19/2014 43.00 | Judy Qiu, Xiaoming Gao. Supporting Queries and Analyses of Large-Scale Social Media Data with Customizable and Scalable Indexing Techniques over NoSQL Databases, Proceedings of the 14th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid 2014). 26-MAY-14, . : , |
| 08/19/2014 45.00 | P. Senin , S. Malinchik. SAX-VSM: Interpretable Time Series Classi?cation UsingSAXandVectorSpaceModel, Proc. of ICDM 2013, Dallas, Texas December 7-10, 2013. 07-DEC-13, . : , |
| 08/19/2014 46.00 | S. Malinchik. Detection of Persuasion Campaigns on TwitterTM by SAX-VSM Technology, Proc. of ICDS 2014, The Eighth International Conference on Digital Society, Barcelona, Spain, March 2014. 23-MAR-14, . : , |
| 08/19/2014 47.00 | Yue Wang, Paul Resnick, Qiaozhu Mei, Cheng Li. ReQ-ReC: High Recall Retrieval with Query Pooling and Interactive Classification, SIGIR '14 The 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, Gold Coast, QLD, Australia — July 06 - 11, 2014. 06-JUL-14, . : , |
| 08/19/2014 48.00 | Shoubin Kong , Qiaozhu Mei , Ling Feng , Fei Ye , Zhe Zhao. Predicting Bursts and Popularity of Hashtags in Real-Time, SIGIR '14 The 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, Gold Coast, QLD, Australia — July 06 - 11, 2014. 06-JUL-14, . : , |
| 08/19/2014 49.00 | Xin Rong, Qiaozhu Mei. Diffusion of innovations revisited: from social network to innovation network, Proceedings of the 22nd ACM international conference on Conference on information & knowledge management, San Francisco, CA, USA — October 27 - November 01, 2013. 27-OCT-13, . : , |
| 08/19/2014 51.00 | Cheng Li, Yue Wang, Qiaozhu Mei. . A User-in-the-Loop Process for Investigational Search: Foreseer in TREC 2013 Microblog Track, . Proceedings of the Twenty-Second Text REtrieval Conference (TREC 2013). 19-NOV-13, . : , |
| 08/19/2015 53.00 | Zhe Zhao, Paul Resnick, Qiaozhu Mei. Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts, 24th international conference on World Wide Web (WWW'15). 18-MAY-15, . : , |
| 08/19/2015 55.00 | Xiaoming Gao , Emilio Ferrara , Judy Qiu. Parallel Clustering of High-Dimensional Social Media Data Streams, 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, 2015. 04-MAY-15, . : , |

08/19/2015 54.00 Cheng Li, Yue Wang, Paul Resnick, Qiaozhu Mei. ReQ-ReC: High-Recall Retrieval with Rate-Limited Queries,
37th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'14). 06-JUL-14, . : ,

08/23/2013 21.00 Emilio Ferrara, Mohsen JafariAsbagh, Onur Varol, Vahed Qazvinian, Filippo Menczer, Alessandro Flammini. Clustering Memes in Social Media,
Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM'13), 2013. IEEE/ACM. 25-AUG-13, . : ,

08/23/2013 22.00 Karissa McKelvey, Filippo Menczer. Design and Prototyping of a Social Media Observatory,
22nd International Conference of Wolrd Wide Web (WWW'13). 13-MAY-13, . : ,

08/23/2013 23.00 Filippo Menczer, Karissa McKelvey. Interoperability of Social Media Observatories,
First International Workshop on Building Web Observatories. 01-MAY-13, . : ,

08/23/2013 24.00 Karissa McKelvey, Filippo Menczer. Truthy: Enabling the study of online social networks,
2013 conference on Computer supported cooperative work (CSCW'13). 23-FEB-13, . : ,

08/23/2013 25.00 L Weng , J Ratkiewicz , N Perra , B Gonçalves , C Castillo, F Bonchi, R Schifanella, F Menczer, A Flammini. The Role of Information Diffusion in the Evolution of Social Networks,
19th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'13). 11-AUG-13, . : ,

08/23/2013 28.00 O Varol, F Menczer, A Flammini, E Ferrara. Traveling Trends: Social Butterflies or Frequent Fliers?,
ACM Conference on Online Social Networks (COSN'13). 08-OCT-13, . : ,

08/23/2013 30.00 Z Zhao, Q Mei. Questions about questions: an empirical analysis of information needs on Twitter,
22nd international conference on World Wide Web (WWW'13). 13-MAY-13, . : ,

**TOTAL: 20**

**Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):**

---

## (d) Manuscripts

Received Paper

08/19/2015 56.00 Emilio Ferrara , Onur Varol , Clayton Davis , Filippo Menczer , Alessandro Flammini. The rise of Social Bots,
Communications of the ACM Journal (06 2015)

**TOTAL: 1**

**Number of Manuscripts:**

## Books

Received Book

**TOTAL:**

Received Book Chapter

08/19/2014 39.00 Xiaoming Gao , Evan Roth , Karissa McKelvey , Clayton Davis , Andrew Younge , Emilio Ferrara , Filippo Menczer , Judy Qiu . Supporting a Social Media Observatory with Customizable Index Structures — Architecture and Performance , : , ( )

08/19/2014 40.00 Evan Roth , Xiaoming Gao , Karissa McKelvey , Clayton Davis , Andrew Younge , Emilio Ferrara , Filippo Menczer , Judy Qiu . Supporting a Social Media Observatory with Customizable Index Structures — Architecture and Performance , New York: Cloud Computing for Data Intensive Applications, Springer, 2014 , (11 2014)

08/19/2015 58.00 Lilian Weng, Filippo Menczer, Alessandro Flammini. Online Interactions, Switzerland: Springer International Publishing, (09 2015)

**TOTAL:** **3**

## Patents Submitted

## Patents Awarded

## Awards

## Graduate Students

| NAME | PERCENT_SUPPORTED | Discipline |
|---|---|---|
| Yang, Zeyao | 0.49 | |
| Varol, Onur | 1.00 | |
| Shiralkar, Prashant | 0.49 | |
| Park, Jaehyuk | 0.49 | |
| Zhao, Zhe | 0.40 | |
| **FTE Equivalent:** | **2.87** | |
| **Total Number:** | **5** | |

## Names of Post Doctorates

| NAME | PERCENT_SUPPORTED |
|---|---|
| Ferrara, Emilio | 0.49 |
| Ciampaglia, Giovanni Luca | 0.32 |
| Vydiaswaran, Vinod VG | 1.00 |
| **FTE Equivalent:** | **1.81** |
| **Total Number:** | **3** |

## Names of Faculty Supported

| NAME | PERCENT_SUPPORTED | National Academy Member |
|---|---|---|
| Qiaozhu Mei | 0.04 | |
| **FTE Equivalent:** | **0.04** | |
| **Total Number:** | **1** | |

## Names of Under Graduate students supported

| NAME | PERCENT_SUPPORTED |
|---|---|
| **FTE Equivalent:** | |
| **Total Number:** | |

## Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: ..... 0.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:..... 0.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):..... 0.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense ..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:..... 0.00

## Names of Personnel receiving masters degrees

NAME

**Total Number:**

## Names of personnel receiving PHDs

NAME

**Total Number:**

## Names of other research staff

NAME PERCENT_SUPPORTED

**FTE Equivalent:**
**Total Number:**

## Sub Contractors (DD882)

## Inventions (DD882)

# Scientific Progress

PLEASE SEE ATTACHMENT FOR FULL REPORT. BELOW LIST OF SECTIONS, FIGURES, TABLES, and REFERENCED PAPERS. ALSO BELOW a SUMMARY of the SCIENTIFIC PROGRESS

List of Sections



List of Tables



List of Figures

NMI values at each step are averaged across five runs. These values are then accumulated over the course of the experiment. The inset plots the time-averaged LFK-NMI, with error bars corresponding to standard error.

Fig. 5.2.2: Overlap (Jaccard coefficient) between ground truth classes and clusters detected by PSC (left), B2 (middle), and B1 (right).

Fig. 5.3.2.1: The tradeoff between intra- and inter-community spreading. Stronger communities (small ) facilitate spreading within the originating community (local) while weak communities (large ) provide bridges that allow spreading between communities (global). There is a range of values that allow both (optimal). The blue squares represents , the final density of active nodes in the community A, and the red circles represents . The parameters for the simulation are: , , N = 131,056 and z = 20.

Fig. 5.3.2.2: (a) the phase diagram of threshold model in the presence of community structures with N = 131,056 and z=20, and . There are three phases: no diffusion (white), local diffusion that saturates the community A (blue), and global diffusion (red). The dotted and dashed lines indicate the values of shown in (b) and (c). (b) the cross-sections of the phase diagram (dotted lines in (a)). TL (solid lines) shows excellent agreements with the simulation while MF (dotted lines) overestimate the possibility of global diffusion. (c) the cross-sections represented in dashed lines in (a).

Fig. 5.4.1.1: (left) Trend similarity matrix for 12 cities in Turkey. From the dendrogram on top we can isolate three distinct clusters. (right) Location of the cities with trend information, labeled by the three clusters induced by trend similarity.

Fig 5.4.1.2: Hourly volume of tweets, retweets and replies between May 30th and June 20th, 2013 (top). The timeline is annotated with events from Table 1 of ref. [6]. User (center) and hashtag (bottom) hourly and cumulative volume of tweets over time.

Fig. 5.4.1.3: (left) Distribution of friends and followers of users involved in the Gezi Park conversation; (right) Distribution of user roles as function of social ties and interactions.

Fig. 5.4.1.4: Average displacement of roles over time for the four different classes of roles. The size of the circles represents the number of individuals in each role.

Fig. 5.4.2.1: Geographical differences in language usage in a grassroot meme (#ows).

Fig. 5.4.2.2: Group formation and connectivity evolution for the users involved in the #ows conversation.

Fig 5.4.2.3. Trendsetters vs. trend-followers: the inset shows a Gaussian Mixture Model showing two different trendsetting dynamics; the contours represent the std. dev. of each Gaussian distribution. The main plot shows their linear regressions.

Fig. 5.4.2.4: Features describing different classes of users. Each box shows data within lower and upper quartile; whiskers represent 99th percentile; the triangle and the line in a box represent median and mean, respectively.

Fig. 5.5.1.1: Classification performance of “Bot or Not?” for four different classifiers. The classification accuracy is computed by 10-fold cross validation and measured by the area under the receiver operating characteristic curve (AUROC). The best score, obtained by Random Forest, is 95%.

Fig. 5.5.1.2: Subset of user features that best discriminate social bots from humans. Bots retweet more than humans and have longer user names, while they produce fewer tweets, replies and mentions, and they are retweeted less than humans. Bot accounts also tend to be more recent.

Fig. 5.5.1.3: Visualizations provided by “Bot or Not?.” (A) Part-of-speech tag proportions. (B) Language distribution of contacts. (C) Network of co-occurring hashtags. (D) Emotion, happiness and arousal-dominance-valence sentiment scores. (E) Temporal patterns of content consumption and production

Fig. 5.5.2.1: Distribution of cosine similarity between pairs of accounts.

Fig. 5.5.2.2: Hashtag co-occurrence networks.

Fig. 5.5.2.3: Visualization and interactive data inspection interface.

Fig. 5.6.1.1: Using Wikipedia to fact-check statements. (a) To populate the knowledge graph with facts we use structured information contained in the ‘info-boxes’ of Wikipedia articles (in the figure, the info-box of the article about Barack Obama). (b)

In the diagram we plot the shortest path returned by our method for the statement “Barack Obama is a Muslim.” The path traverses high-degree nodes representing generic entities, such as Canada, and is assigned a low truth-value.

Fig. 5.6.3.1: Automatic truth assessments for simple factual statements. In each confusion matrix, rows represent subjects and columns represent objects. The diagonals represent true statements. Higher truth-values are mapped to colors of increasing intensity. (a) Films winning the Oscar for Best Movie and their directors, grouped by decade of award. (b) US presidents and their spouses, denoted by initials. (c) US states and their capitals, grouped by US Census Bureau-designated regions. (d) World countries and their capitals, grouped by continent.

Fig. 5.7.1.1: A visualization of the SAX dimensionality reduction technique.
A time series (red line) is discretized first by a PAA procedure (N = 8) and then, using breakpoints of arbitrary length, it is mapped into the word “baabccbc” using an alphabet size of 3.

Figure 5.7.1.2: By sliding a window across time series, extracting subsequences, converting them to SAX words, and placing these words into an unordered collection, we obtain the bag of words representation of the original time series. Next, TF*IDF statistics is computed resulting in a single weight vector.

Figure 5.7.1.3: An overview of SAX-VSM algorithm: at first, all labeled time series from each class are converted into a single bag of words using SAX; secondly, TF*IDF statistics is computed resulting in a single weight vector per training class. For classification, an unlabeled time series is converted into a term frequency vector and assigned a label of a weight vector, which yields a maximal cosine similarity value.

Figure 5.7.1.4: SAX-VSM algorithm characteristics: (a) - Comparison of classification precision and run time of SAX-VSM (red) and 1NN Euclidean classifier (blue) on CBF data. SAX-VSM performs significantly better with limited amount of training samples. (b) - While SAX-VSM is faster in time series classification, its performance is comparable to 1NN Euclidean classifier when training time is accounted for. (c) -SAX-VSM increasingly outperforms 1NN Euclidean classifier with the growth of a noise level.

Figure 5.7.2.1: ROC curve for the classification experiment

Figure 5.7.3.1. Improving detection accuracy during the feature selection process.

Figure 5.8.2.1: Longitudinal analysis showing examples of tweets with information need [2].

Figure 5.8.3.1: The procedure of real-time rumor detection (Figure 1 in Zhao et al. [3]).

Figure 5.8.3.2: Detect and track rumors rom Boston marathon explosion (Figure 5 in Zhao et al. [3]).

Figure 5.8.4.1: ReQ-ReC framework (Figure 1 in Li et al. [4]).

List of References

1. Varol, F Menczer: Connecting Dream Networks Across Cultures. WWW Companion '14: Proceedings of the companion publication of the 23rd international conference on World Wide Web companion, 2014. http://dx.doi.org/10.1145/2567948.2579697¿
2. Xiaoming Gao, Evan Roth, Karissa McKelvey, Clayton Davis, Andrew Younge, Emilio Ferrara, Filippo Menczer, Judy Qiu: Supporting a Social Media Observatory with Customizable Index Structures-Architecture and Performance. Book chapter in Cloud Computing for Data Intensive Applications, Springer, 2014 http://salsaproj.indiana.edu/IndexedHBase/paper_bookChapter.pdf¿
3. A Nematzadeh, E Ferrara, A Flammini, and YY Ahn. Optimal network clustering for information diffusion. Physical Review Letters, 113, 088701, 2014. [Editor’s pick] http://journals.aps.org/prl/abstract/10.1103/PhysRevLett.113.088701¿
4. M JafariAsbagh, E Ferrara, O Varol, F Menczer, and A Flammini. Clustering memes in social media streams. Social Network Analysis and Mining Social Network Analysis and Mining 4 (237), 1-13, 2014. http://link.springer.com/article/10.1007/s13278-014-0237-x#page-1

5. E Ferrara, O Varol, C Davis, F Menczer, A Flammini. The rise of Social Bots. Communications of the ACM – (to appear) http://arxiv.org/abs/1407.5225¿
6. Varol, E Ferrara, C Ogan, F Menczer, and A Flammini. Evolution of online user behavior during a social upheaval. ACM Web Science ’14, pp. 81-90. ACM 2014 [Best Paper Award] http://dl.acm.org/citation.cfm?id=2615699¿
7. Xiaoming Gao, Judy Qiu. Supporting Queries and Analyses of Large-Scale Social Media Data with Customizable and Scalable Indexing Techniques over NoSQL Databases. Proceedings of the 14th IEEE/ACM International Symposium on

Cluster, Cloud and Grid Computing (CCGrid 2014). Chicago, IL, USA, May 26-29, 2014 http://mypage.iu.edu/~gao4/paper_ccgrid2014.pdf¿
8. Xiaoming Gao, Judy Qiu. Social Media Data Analysis with IndexedHBase and Iterative MapReduce. Proceedings of the 6th Workshop on Many-Task Computing on Clouds, Grids, and Supercomputers Clouds, Grids, and Supercomputers (MTAGS 2013) at Super Computing 2013. Denver, CO, USA, November 17th, 2013. http://datasys.cs.iit.edu/events/MTAGS13/p07.pdf

1. Lilian Weng, Filippo Menczer. Topicality and social impact: diverse messages but focused messangers. PLoS One, 10(2): e0118410, 2015 http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0118410 ¿

2. P. Senin and S. Malinchik, SAX-VSM: Interpretable Time Series Classification Using SAX and Vector Space Model, Proc. of ICDM 2013, Dallas, Texas / December 7-10, 2013
http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=6729617¿

3. S. Malinchik, Detection of Persuasion Campaigns on TwitterTM by SAX-VSM Technology, Proc. of ICDS 2014, The Eighth International Conference on Digital Society, Barcelona, Spain, March 2014.
http://www.thinkmind.org/index.php?view=article&articleid=icds_2014_5_20_10080
¿
4. Cheng Li, Yue Wang, Paul Resnick, Qiaozhu Mei. ReQ-ReC: High Recall Retrieval with Query Pooling and Interactive Classification. SIGIR '14 The 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, Gold Coast, QLD, Australia — July 06 - 11, 2014 http://www-personal.umich.edu/~qmei/pub/sigir2014-li.pdf¿

5. Shoubin Kong, Qiaozhu Mei, Ling Feng, Fei Ye, Zhe Zhao. Predicting Bursts and Popularity of Hashtags in Real-Time. SIGIR '14 The 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, Gold Coast, QLD, Australia — July 06 - 11, 2014 http://www-personal.umich.edu/~qmei/pub/sigir2014-kong.pdf¿

6. Xin Rong, Qiaozhu Mei. Diffusion of innovations revisited: from social network to innovation network. Proceedings of the 22nd ACM international conference on Conference on information & knowledge management, San Francisco, CA, USA — October 27 - November 01, 2013 http://www-personal.umich.edu/~qmei/pub/kdd2013-tang.pdf¿

7. Zhe Zhao, Paul Resnick and Qiaozhu Mei, Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts, WWW '15 Proceedings of the 24th International Conference on World Wide Web, 2015 http://dl.acm.org/citation.cfm?id=2741637 ¿

8. Cheng Li, Yue Wang, and Qiaozhu Mei, A User-in-the-Loop Process for Investigational Search: Foreseer in TREC 2013 Microblog Track, in Proceedings of the Twenty-Second Text REtrieval Conference (TREC 2013). http://trec.nist.gov/pubs/trec22/papers/foreseer-microblog.pdf¿

9. Ferrara, E., JafariAsbagh, M., Varol, O., Qazvinian, V., Menczer, F., & Flammini, A. Clustering Memes in Social Media. In: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM' 13), 2013. IEEE/ACM
http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6785757

10. McKelvey, K., & Menczer, F. Design and prototyping of a social media observatory. In: Proceedings of the 22nd international conference on World Wide Web companion, pp. 1351-1358, May 13–17, 2013, Rio de Janeiro, Brazil. ACM 2013. http://dl.acm.org/citation.cfm?id=2488174

11. McKelvey, K., & Menczer, F. Interoperability of Social Media Observatories. In: Proceedings of the First International Workshop on Building Web Observatories. May 8, 2013, Paris, France. ACM http://cnets.indiana.edu/wp-content/uploads/websci13.pdf

12. McKelvey, K. R., & Menczer, F. Truthy: enabling the study of online social networks. In: Proceedings of the 2013 conference on Computer supported cooperative work companion, pp. 23-26, 2013. ACM 2013. http://dl.acm.org/citation.cfm?id=2441962

13. Weng, L., Ratkiewicz, J., Perra, N., Gonçalves, B., Castillo, C., Bonchi, F., Schifanella, R., Menczer, F., & Flammini, A. (2013). The Role of Information Diffusion in the Evolution of Social Networks. In: Proceedings of the 19th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), August 11-14, 2013, Chicago, USA. http://dl.acm.org/citation.cfm?id=2487607

14. Conover, M. D., Davis, C., Ferrara, E., McKelvey, K., Menczer, F., & Flammini, A. (2013). The geospatial characteristics of a social movement communication network. PloS one, 8(3), e55957
http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0055957

15. Conover, M. D., Ferrara, E., Menczer, F., & Flammini, A. (2013). The Digital Evolution of Occupy Wall Street. PloS one, 8 (5), e64679. http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0064679

16. Ferrara, E., Varol, O., Menczer, F., & Flammini, A. Traveling Trends: Social Butterflies or Frequent Fliers? In: Proceedings of the ACM Conference on Online Social Networks (COSN 2013), 2013. ACM 978-1-4503-2084-9/13/10 http://dl.acm.org/citation.cfm?id=2512956

17. Zhe Zhao and Qiaozhu Mei, "Questions about questions: an empirical analysis of information needs on Twitter," in Proceedings of the 22nd international conference on World Wide Web (WWW'13), pp. 1545-1556, 2013. http://dl.acm.org/citation.cfm?id=2488523

18. X Gao, E Ferrara, J Qiu. Parallel Clustering of High-Dimensional Social Media Data Streams. 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computin, 2015. http://arxiv.org/abs/1502.00316

19. Samuel Carton, Souneil Park, Nicole Zeffer, Eytan Adar, Qiaozhu Mei, and Paul Resnick, “Audience Analysis for Competing Memes in Social Media,” ICWSM 2015. http://www.cond.org/rumorlens_icwsm_2015_final.pdf

20. VS Subrahmanian, O Varol, P Shiralkar, E Ferrara, F Menczer, A Flammini, et al. The DARPA Twitter Bot Challenge. IEEE Computer (to appear), 2015.

21. Weng, L., Ratkiewicz, J., Perra, N., Gonçalves, B., Castillo, C., Bonchi, F., Schifanella, R., Menczer, F., & Flammini, A. (2013). The Role of Information Diffusion in the Evolution of Social Networks. In: Proceedings of the 19th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), August 11-14, 2013, Chicago, USA. http://dl.acm.org/citation.cfm?id=2487607

22. Conover, M. D., Davis, C., Ferrara, E., McKelvey, K., Menczer, F., & Flammini, A. (2013). The geospatial characteristics of a social movement communication network. PloS one, 8(3), e55957 http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0055957

23. Conover, M. D., Ferrara, E., Menczer, F., & Flammini, A. (2013). The Digital Evolution of Occupy Wall Street. PloS one, 8 (5), e64679. http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0064679

24. Ferrara, E., Varol, O., Menczer, F., & Flammini, A. Traveling Trends: Social Butterflies or Frequent Fliers? In: Proceedings of the ACM Conference on Online Social Networks (COSN 2013), 2013. ACM 978-1-4503-2084-9/13/10 http://dl.acm.org/citation.cfm?id=2512956

25. Zhe Zhao and Qiaozhu Mei, "Questions about questions: an empirical analysis of information needs on Twitter," in Proceedings of the 22nd international conference on World Wide Web (WWW'13), pp. 1545-1556, 2013. http://dl.acm.org/citation.cfm?id=2488523

26. X Gao, E Ferrara, J Qiu. Parallel Clustering of High-Dimensional Social Media Data Streams. 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computin, 2015. http://arxiv.org/abs/1502.00316

27. Samuel Carton, Souneil Park, Nicole Zeffer, Eytan Adar, Qiaozhu Mei, and Paul Resnick, “Audience Analysis for Competing Memes in Social Media,” ICWSM 2015. http://www.cond.org/rumorlens_icwsm_2015_final.pdf

28. VS Subrahmanian, O Varol, P Shiralkar, E Ferrara, F Menczer, A Flammini, et al. The DARPA Twitter Bot Challenge. IEEE Computer (to appear), 2015.

29. GL Ciampaglia, P Shiralkar, LM Rocha, J Bollen, F Menczer, A Flammini. (2015) Computational Fact Checking from Knowledge Networks. PLoS ONE 10(6):e0128193. http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0128193

SUMMARY

The goal of the DESPIC project was to develop a technological infrastructure to automatically detect orchestrated campaigns in their early stage of diffusion. The list of such campaigns includes rumors, spread of misinformation, persuasion attempts, and advertising. During the performance period we advanced in a number of directions relevant to the stated goal. These include the design and implementation of a high performance infrastructure for data collection, filtering and analysis, and the design and development of the technical and algorithmic framework for detection of (simple) campaigns-the details are outlined below. Our work lead to a number of prestigious publications appeared in top journals in physics and computer science, including

Physical Review Letters, Communications of the ACM, IEEE Computer, PLoS One, and major computer science conferences such as ACM WWW, ACM SIGIR, ACM Web Science, ACM/IEEE ASONAM, and ICWSM, to mention a few.
Our team has received recognition in various forms: the Indiana team became part of the Web Observatory, a group of 8 international teams that includes Oxford, Rensselaer Polytechnic Institute, Southampton University, and KAIST, whose goal is to design frameworks for real-time analysis of activity on social media. We also freely delivered a proof-of-concept of our social bot detection system, called 'Bot or Not?' that allows to automatically classify any Twitter account as human- or bot-like. Our work has attracted attention and been reported about in technology magazines. The list includes the MIT Technology Review, Scientific American, Wired, the Smithsonian Magazine, Vice and Mashable, and in news media outlet like the New York Times, the Washington Post, the Guardian, Politico and the BBC. The initial version of the ReQ-ReC framework discussed below was used to participate in the microblog track of the TREC 2013 and achieved the top rankings among more than 70 submissions from 20 participating teams [16]. A case study on user roles in online discussion and how they change over time [6] received the best paper award at the ACM Websci 2014 conference.

Below a short summary of the main results achieved:
1. The IU Team designed and implemented a distributed infrastructure for Twitter data collection, archival and retrieval. After testing different distributed NoSQL databases including HBase and Riak, we finalized the requirements of the optimal computational architecture to support our framework. We also concluded the development of an API to allow general public to access our datasets in aggregated form. Details about this task are in Sec. 5.1 and published in Cloud Computing for Data Intensive Applications and presented in two prestigious HPC conferences.

2. The IU Team completed the implementation and performance evaluation of an algorithm for tweet clustering in a streaming scenario. The results, despite the limited ground-truth data used for validation, are extremely satisfactory. The system proved able to outperform various baseline methods and state-of-the-art algorithms developed by other groups. Our strategy compensates the paucity of text in Twitter messages by leveraging network, diffusion, and user meta-data to assess tweet similarity. Performance evaluation determined that our framework is able to successfully retrieve trending topics. Further testing will be required to establish whether the topical clustering works also with other conversations of general interest. This research line is discussed in details in Sec. 5.2 and in revision on Social Network Analysis and Mining.

3. To understand the diffusion of misinformation and campaigns on a more theoretical ground, the IU Team investigated how social reinforcement and the modular structure of the social network affect information diffusion. Our findings show that there exists an optimal modular structure that simultaneously enhances local diffusion and global spreading. Our work has been published on the prestigious Physical Review Letters and it was selected as the issue Editor's Pick. See Sec. 5.3.

4. One of the main goals of our project is to understand user behavior on social media platforms. IU Team performed different several case studies, focused on: (i) the geographical and temporal dynamics of the Occupy Wall Street protest. Results are published in two highly-cited papers both in PLoS One. (ii) The Gezi Parki protest, a social upheaval unfolded in Turkey in 2013. The studied examined how user influence and user roles change over time. Our findings, discussed in Sec. 5.4, have been presented at the prestigious ACM Web Science 2014 Conference and the paper received the Conference Best Paper Award. (iii) The evolution of the social network supported by the Yahoo! Meme platform. The goal of the study was to understand the feedback influence loop between information diffusion and the creation of new social links. Our results have been presented at the prestigious ACM KDD 2013 Conference.

5. Social bots are tools increasingly adopted to sustain promotion and persuasion campaigns. IU Team developed a machine-learning framework for the classification of Twitter accounts that discriminates between humans and social bots. Such system was one of the tools we used to approach the DARPA Bot Detection Challenge. Our framework achieves a classification performance above 95% as measured by ROC AUC. The system has been implemented as a open-access platform called 'Bot or Not?' (truthy.indiana.edu/botornot) and has received coverage by several major international news outlets. Our results are described in details in Sec. 5.5 and are appearing under the prestigious Communications of ACM. The performance at the DARPA Challenge will appear on IEEE Computer.

6. Coordinated efforts to promote content in a less than transparent fashion goes often hand-in-hand with the spread of unreliable information. IU studied the possibility to check the validity of simple statements in an automatic fashion. Our framework infers their plausibility of such statements on the basis of the proximity (in a given knowledge graph) of the entities it involves. We found that the complexities of human fact-checking can be approximately resolved by finding the shortest path between concept nodes under properly defined semantic proximity metrics on knowledge graphs. Framed as a network problem, our approach is feasible with efficient computational techniques. We evaluated this approach by examining thousands of claims related to history, entertainment, geography, and biographical information using a public knowledge graph extracted from Wikipedia. Statements independently known to be true consistently receive higher support via our method than do false ones. See Sec. 5.6.

7. The LM ATL Team, in collaboration with the IU team focused on: (a) Exploration of performance of ATL-developed SAX-VSM technology and its ability to detect promoted topics before their trending phase on Twitter; (b) Improving detection

performance of the method by optimizing the feature selection process; (c) Developing and implementing a general approach to detect any type of pattern associated with anomalous information diffusion on Twitter. These include patterns associated to coordinated effort of promotion and persuasion, as well as grass root conversations. See Sec. 5.7.

8. The UM Team developed a real-time algorithm that is able to detect emerging rumors from the stream of social media at a high precision and hours earlier than existing methods. We developed a user-in-the-loop retrieval system that aim to find all posts related to a given rumor, which yielded a 20% to 30% improvement over the state-of-the-art and won the microblog track of the annual TREC competition. We explored the prediction of the burstiness and popularity of hashtags at various stages of their life cycle. We identified and tested the effectiveness of seven types of features. See Sec. 5.8.

## Technology Transfer

**Final Report for:**
**DESPIC: Detecting Early Signatures of Persuasion in Information Cascades**

Program: DARPA SMISC
Period of Performance: January 18th, 2012 – May 31st, 2015
Authors: Alessandro Flammini, Filippo Menczer (Indiana University), Qiaozhu Mei (University of Michigan), Sergey Malinchik (Lockheed Martin Advanced Technology Laboratory)

## 1. Foreword and Extended Summary: Goals, Methods and Results

The goal of the DESPIC project was to develop a technological infrastructure to automatically detect orchestrated campaigns in their early stage of diffusion. The list of such campaigns includes rumors, spread of misinformation, persuasion attempts, and advertising. During the performance period we advanced in a number of directions relevant to the stated goal. These include the design and implementation of a high performance infrastructure for data collection, filtering and analysis, and the design and development of the technical and algorithmic framework for detection of (simple) campaigns-the details are outlined below.

Our work lead to a number of prestigious publications appeared in top journals in physics and computer science, including Physical Review Letters, Communications of the ACM, IEEE Computer, PLoS One, and major computer science conferences such as ACM WWW, ACM SIGIR, ACM Web Science, ACM/IEEE ASONAM, and ICWSM, to mention a few.

Our team has received recognition in various forms: the Indiana team became part of the Web Observatory, a group of 8 international teams that includes Oxford, Rensselaer Polytechnic Institute, Southampton University, and KAIST, whose goal is to design frameworks for real-time analysis of activity on social media. We also freely delivered a proof-of-concept of our social bot detection system, called 'Bot or Not?' that allows to automatically classify any Twitter account as human- or bot-like. Our work has attracted attention and been reported about in technology magazines. The list includes the MIT Technology Review, Scientific American, Wired, the Smithsonian Magazine, Vice and Mashable, and in news media outlet like the New York Times, the Washington Post, the Guardian, Politico and the BBC. The initial version of the ReQ-ReC framework discussed below was used to participate in the microblog track of the TREC 2013 and achieved the top rankings among more than 70 submissions from 20 participating teams [16]. A case study on user roles in online discussion and how they change over time [6] received the best paper award at the ACM Websci 2014 conference.

Below a short summary of the main results achieved:

1. The IU Team designed and implemented a distributed infrastructure for Twitter data collection, archival and retrieval. After testing different distributed NoSQL databases including HBase and Riak, we finalized the requirements of the optimal

computational architecture to support our framework. We also concluded the development of an API to allow general public to access our datasets in aggregated form. Details about this task are in Sec. 5.1 and published in *Cloud Computing for Data Intensive Applications* and presented in two prestigious HPC conferences.

2. The IU Team completed the implementation and performance evaluation of an algorithm for tweet clustering in a streaming scenario. The results, despite the limited ground-truth data used for validation, are extremely satisfactory. The system proved able to outperform various baseline methods and state-of-the-art algorithms developed by other groups. Our strategy compensates the paucity of text in Twitter messages by leveraging network, diffusion, and user meta-data to assess tweet similarity. Performance evaluation determined that our framework is able to successfully retrieve trending topics. Further testing will be required to establish whether the topical clustering works also with other conversations of general interest. This research line is discussed in details in Sec. 5.2 and in revision on *Social Network Analysis and Mining*.

3. To understand the diffusion of misinformation and campaigns on a more theoretical ground, the IU Team investigated how social reinforcement and the modular structure of the social network affect information diffusion. Our findings show that there exists an optimal modular structure that simultaneously enhances local diffusion and global spreading. Our work has been published on the prestigious *Physical Review Letters* and it was selected as the issue Editor's Pick. See Sec. 5.3.

4. One of the main goals of our project is to understand user behavior on social media platforms. IU Team performed different several case studies, focused on: (i) the geographical and temporal dynamics of the Occupy Wall Street protest. Results are published in two highly-cited papers both in PLoS One. (ii) The Gezi Parki protest, a social upheaval unfolded in Turkey in 2013. The studied examined how user influence and user roles change over time. Our findings, discussed in Sec. 5.4, have been presented at the prestigious *ACM Web Science 2014 Conference* and the paper received the *Conference Best Paper Award*. (iii) The evolution of the social network supported by the Yahoo! Meme platform. The goal of the study was to understand the feedback influence loop between information diffusion and the creation of new social links. Our results have been presented at the prestigious *ACM KDD 2013 Conference*.

5. Social bots are tools increasingly adopted to sustain promotion and persuasion campaigns. IU Team developed a machine-learning framework for the classification of Twitter accounts that discriminates between humans and social bots. Such system was one of the tools we used to approach the DARPA Bot Detection Challenge. Our framework achieves a classification performance above 95% as measured by ROC AUC. The system has been implemented as a open-access platform called 'Bot or Not?' (truthy.indiana.edu/botornot) and has received coverage by several major international news outlets. Our results are described in details in Sec. 5.5 and are appearing under the prestigious *Communications of ACM*. The performance at the DARPA Challenge will appear on *IEEE Computer*.

6. Coordinated efforts to promote content in a less than transparent fashion goes often hand-in-hand with the spread of unreliable information. IU studied the possibility to check the validity of simple statements in an automatic fashion. Our framework infers their plausibility of such statements on the basis of the proximity (in a given knowledge graph) of the entities it involves. We found that the complexities of human fact-checking can be approximately resolved by finding the shortest path between concept nodes under properly defined semantic proximity metrics on knowledge graphs. Framed as a network problem, our approach is feasible with efficient computational techniques. We evaluated this approach by examining thousands of claims related to history, entertainment, geography, and biographical information using a public knowledge graph extracted from Wikipedia. Statements independently known to be true consistently receive higher support via our method than do false ones. See Sec. 5.6.

7. The LM ATL Team, in collaboration with the IU team focused on: (a) Exploration of performance of ATL-developed SAX-VSM technology and its ability to detect promoted topics before their trending phase on Twitter; (b) Improving detection performance of the method by optimizing the feature selection process; (c) Developing and implementing a general approach to detect any type of pattern associated with anomalous information diffusion on Twitter. These include patterns associated to coordinated effort of promotion and persuasion, as well as grass root conversations. See Sec. 5.7.

8. The UM Team developed a real-time algorithm that is able to detect emerging rumors from the stream of social media at a high precision and hours earlier than existing methods. We developed a user-in-the-loop retrieval system that aim to find all posts related to a given rumor, which yielded a 20% to 30% improvement over the state-of-the-art and won the microblog track of the annual TREC competition. We explored the prediction of the burstiness and popularity of hashtags at various stages of their life cycle. We identified and tested the effectiveness of seven types of features. See Sec. 5.8.

## 2. Team members

IU: Alessandro Flammini (Team PI), Filippo Menczer (co-PI)
UM: Qiaozhu Mei (co-PI)
LM ATL: Sergey Malinchik (co-PI)

## 3. Students & Postdocs

- Both supported, partially supported, or involved but not supported
- Info below refers to period of performance Aug. 1st, 2014 to May 31st, 2015

Emilio Ferrara – PostDoc – IU
V.G. Vinod Vydiswaran– Postdoc – UM
Giovanni Ciampaglia – Postdoc - IU

Onur Varol - PhD student – IU
Xioaming Gao – PhD student - IU
Prashant Shiralkar – Phd student – IU
Zhe Zhao – PhD student – UM
Zheyao Yang – PhD student – IU
Jaeyuk Park – PhD student - IU
Vaishnav Kameswaran – Master student - UM
Xiaochen Li – Master student - UM

## 4. Meetings

- F. Menczer (Co-PI, IU), A. Flammini (PI, IU), E. Ferrara, S. Malinchik (co-PI, LM ATL) and Q. Mei (co-PI, UMich) participated to the bi-annual DARPA SMISC-ADAMS meetings held in Arlington, VA in Oct. 2013 and Apr. 2014
- F. Menczer, A. Flammini, E. Ferrara, and S. Malinchik met for project internal review at Lockheed Martin offices in Arlington, VA in Apr. 2014
- All teams participated in monthly conference calls for updates and project coordination.
- F. Menczer and E. Ferrara joined several teleconferences to discuss the design and delivery of the DARPA SMISC Challenge on the detection of synthetic accounts on Twitter.
- F. Menczer and E. Ferrara have been participating in the Data working group teleconference calls
- A. Flammini and E. Ferrara have been participating in the System working group teleconference calls
- S. Malinchik have been participating in the Metrics working group
- Q. Mei participated in the Test & Evaluation working group
- Member of the team presented the results of the work presented here in a consistent number of conferences and workshops. Their list is reported in the appropriate box of the reporting website

## 5. Research Activities and Results

### 5.1 Distributed infrastructure and Data Analysis API development

In order to store the data set retrieved from the Twitter streaming API and support various data queries and analysis tasks, the IU Team developed a scalable data infrastructure. This infrastructure is based on Apache YARN (Hadoop 2) and HBase, and has been deployed on the “Moe” cluster hosted at Indiana University. The hardware configuration of the cluster is given in Table 5.1.1.

Fig. 5.1.1 illustrates the architecture of this data infrastructure, featuring the following advantages:

(1) Fine-grained data access to each social update (tweet) and its associated user information. This forms the basis for efficient query evaluation and analysis algorithm execution.
(2) A novel customizable indexing framework [2,7,8] to build the most suitable index structures for query and analysis purposes. Fig. 5.1.2 gives an example of customized index structure that cannot be constructed by using existing text indexing systems such as Lucene.
(3) Efficient parallel data loading and indexing strategies for both static and streaming data.
(4) Dynamic adoption of different parallel computing frameworks (e.g., Hadoop MapReduce, iterative MapReduce, Giraph graph processing, etc.) for different query and post-query analysis tasks. General queries and analysis algorithms can be developed as building blocks for constructing various analysis workflows.

Our previous research has demonstrated that, based on specially customized index structures, we can achieve much more efficient data loading, query evaluation and analysis task solutions compared with using traditional methods. Table 5.1.2 compares the data loading performance between our solution based on Hadoop MapReduce over HBase and another implementation based on Riak, a widely adopted commercial NoSQL database. At its backend, Riak uses distributed Solr, the de facto text indexing technology in industry. Thanks to the customized index structures, as well as better data normalization and compression with HBase, we can achieve both smaller loaded data sizes and a 6 times faster data loading speed than the Riak-based implementation. Fig. 5.1.3 shows the scalable performance of our parallel stream data loading strategy in an 8-node testing environment. With 8 distributed loaders, our strategy can load one day's data (in the example, 2013-07-13) within less than 4 hours. This means that the maximum stream data speed that can be handled is 5 times faster than the data rate of one day's data.

Fig. 5.1.4 compares the performance of our parallel query evaluation strategy (IndexedHBase) against two other implementations – one raw data scan solution using Hadoop MapReduce (Hadoop-FS), and another implementation using the text indexing and MapReduce mechanisms on Riak, with an example query get-mention-edges (#euro2012, [2012-06-01, 2012-06-30]). Our solution is 10s to 100s of times faster than the raw data scan solution, and multiple times faster than using Riak. Fig. 5.1.5 further demonstrates that our post-query analysis algorithms using customized index structures are also 10s of times faster than raw data scan solutions for two typical analysis tasks: related hashtag mining and daily meme frequency generation. The details of our work are reported in few papers [2,7,8] appeared in top information systems conferences.

At the same time the IU Team has worked on the development of the Truthy API[1] that provides a simple interface for accessing our data and statistics using simple scripts. We currently offer our application-programming interface (API) through Mashape[2] that provides modules in various languages (Java, PHP, Python, Ruby, Obj-C), used to access the API in JSON, CSV, or XML format. The goal of this tool is to provide a user-friendly framework to analyze social media data accessible to a large public of scholars in different disciplines.

| Node type | # of nodes | Software role | CPU | RAM | Hard Disk | Network |
|---|---|---|---|---|---|---|
| Head node | 3 | HDFS Name node, YARN resource mananger, Zoo-keepers, HBase master | 2 * Intel 6-core E 5-2620v2 | 64 GB | 240GB SSD, 4TB SATA HDD (shared) | 10Gb Ethernet |
| Compute node | 10 | HDFS data nodes, YARN node managers, HBase region servers | 2 * Intel 8-core E5-2650v2 | 128GB | 240GB SSD , 48TB SATA HDD | 10Gb Ethernet |

**Table 5.1.1**: Hardware configuration for the Truthy data infrastructure

| | ding time (hours) | ded total data size (GB) | led original data size (GB) | ded index data size (GB) |
|---|---|---|---|---|
| Riak | 294.11 | 3258 | 2591 | 667 |
| IndexedHBase | 45.47 | 1167 | 955 | 212 |
| Riak / IndexedHBase | 6.47 | 2.79 | 2.71 | 3.15 |

**Table 5.1.2**: Historical data loading performance comparison for 2012-06 (352GB)

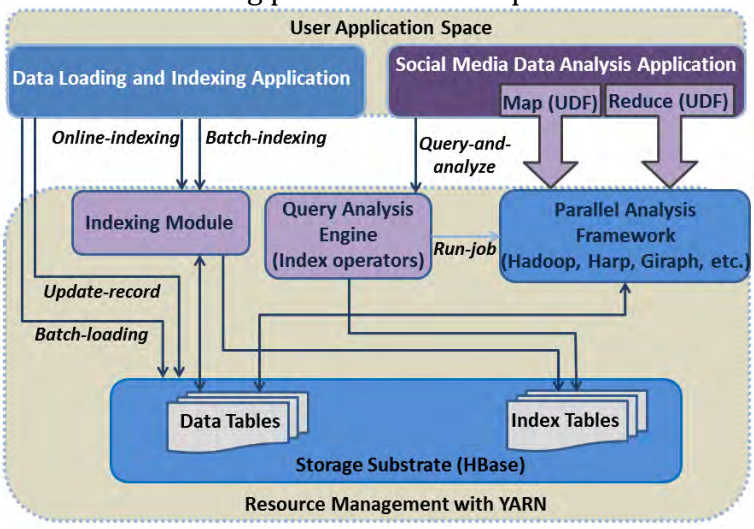



**Fig. 5.1.1**: Architecture of the Truthy data infrastructure

[1] http://truthy.indiana.edu/apidoc
[2] https://www.mashape.com/truthy/truthy-1#!documentation

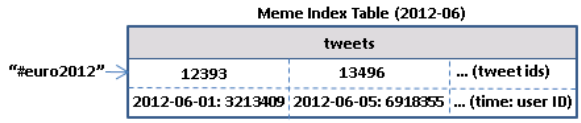



**Fig. 5.1.2**: An example customized index structure

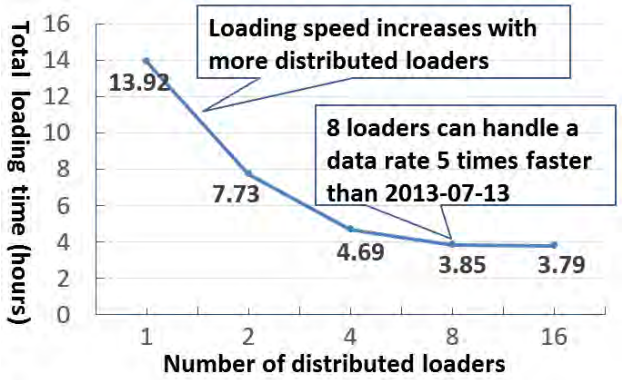



**Fig. 5.1.3**: Scalable parallel stream data loading performance

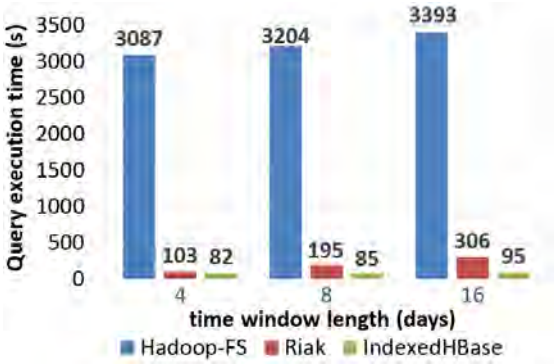



**Fig. 5.1.4**: Query evaluation performance comparison

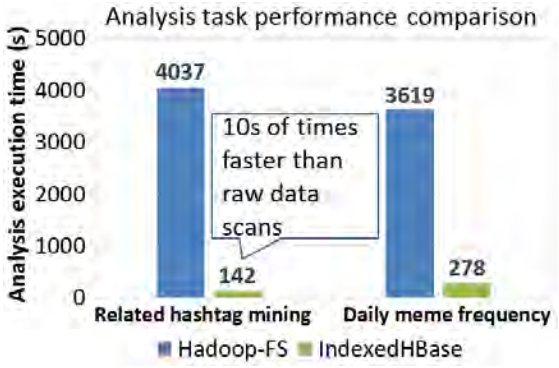



**Fig. 5.1.5**: Analysis task performance comparison

## 5.2 Topical meme clustering and stream clustering

The IU Team has developed a topical meme-clustering framework that works both in static and in streaming scenarios. The underlying common idea to use specific set of tweets (proto-memes) instead of single tweets as basic entities to be clustered. A proto-meme is here defined as a set of tweets carrying a common token of information, such as a hashtag, a URL or a portion of text. The clear advantage of this approach is that each tweet, in principle, can exhibit multiple proto-memes and therefore tweets can carry multiple "concepts" therein. At the same time, the use of protomemes obviates to the scarcity of text of single tweets that hinder the performance of traditional text-based clustering strategies.

Details about the definition of protomemes, the problem of static protomeme clustering, (including the definition of the several protomeme similarity measures we adopted for clustering purpose), and the performance evaluation (including the measure and the dataset employed) are discussed at length in ref. [4].

We here discuss in more detail the novelty introduced during the third year of the DESPIC project to face the problem of real-time clustering in social streams. The model assumption in data stream clustering is that – due to the large amount of incoming data – the system cannot store all of it in memory. Additionally, as a data stream evolves with time, patterns in recent data become more relevant for the clustering algorithm than those in older data. An established way to de-emphasize older data is to represent the stream trough a sliding window-based model that at any time T considers only the last $\ell$ time steps.

*Online K-means* is a simple data stream clustering algorithm based on iterative K-means for stationary data. In general, Online K-means starts with $K$ randomly chosen initial cluster seeds and every new point arriving in the stream is assigned to the closest existing cluster. The closest cluster is chosen based on the distance between the arriving point and the centroid of the cluster. A cluster centroid is described by the same set of coordinates as the data points and the specific coordinates of the centroid are the average of the corresponding coordinates across the data points that are

members of the cluster. This general algorithm can hardly account for new concepts (to be represented by new centroids) might appear in the stream. These new concepts should be represented new clusters; assigning their tweets to existing clusters might jeopardize the quality of clustering. To overcome this problem, one suggested approach is to check whether the distance from the closest cluster centroid is an outlier in comparison to the other closest distances that have been observed so far. If not, the new data point is added to nearest cluster. Otherwise, a new cluster replaces the least recently updated cluster with the new point as the only member. The least recently updated cluster is the one to which no new points have been assigned for the longest time. The outlier detection function uses a history of closest distances from previously observed data points to determine whether a given distance is an outlier. Every time a data point arrives in the stream, its distance to the closest centroid is added to the list. This method assumes that the distances follow a normal distribution. If the new distance exceeds the historical average by $n$ standard deviations or more, where $n$ is a parameter, the new point is deemed an outlier. The proposed clustering algorithm, that we call *Protomeme Stream Clustering* (PSC), works as follows:

1. At the beginning of each step, the sliding window is advanced by $\Delta t$ and protomemes are extracted from arriving tweets in the stream, i.e., those with timestamp in $(T - \Delta t, T]$. Each protomeme is treated as a data point to be clustered. Before these new points are assigned to clusters, all clusters are examined and data points with time stamps older than $T - \ell\Delta t$ (i.e., those that are no longer in the sliding window) are removed. From now on, we will refer to these points as *old* or *expired*. If a cluster consists only of old points, it becomes empty and is removed from the list of clusters.
2. Since we are using protomemes as a pre-aggregation step, in our algorithm we tend to assign the same protomemes to the same clusters whenever possible. If an arriving protomeme matches any of the ones present in any of the existing clusters, we assign it to that cluster and continue to the next protomeme. Otherwise, we move to the next step.
3. A new protomeme is assigned to the closest cluster or to a new cluster based on the outcome of the outlier test. The protomeme is assigned to a new cluster if its distance from the nearest centroid $d > \mu + n\sigma$, where $\mu$ and $\sigma$ are the mean and standard deviation, respectively, of the values in the historical list of closest distance values. The historical distance values in the list are kept since the beginning of the clustering process.

Fig. 5.2.1 plots cumulative LFK-NMI over all the evaluation periods. Each point on the $x$-axis represents a six-hour sliding window terminating at the indicated hour. To compute LFK-NMI correctly for each evaluation period, it is essential to have the same set of tweets in the ground truth and evaluated clusters. Therefore, we only use tweets and their labels in the ground truth for the same period of time. As explained earlier, whenever a cluster becomes empty after removing old data points, we remove it from the list of clusters. In a real world scenario, we might decide to ignore these clusters because they have not been updated during the last $\ell$ time steps; for evaluation purposes we keep them in a separate list and account for them when assessing the

quality in the present window, then delete them afterwards. Our algorithm performs consistently better than the two baselines we considered for comparison. The performance improvement is more apparent when the online clustering has been [3]performed over a sufficiently long period of time. Fig. 5.2.1 shows that after about half of the running time, PSC provides a consistent improvement in cluster quality with respect to the baselines. This is due to the characteristic fast-paced churning time of the topics of discussion in social media. The inset of Fig. 5.2.1 demonstrates that the differences in LFK-NMI between PSC and the baseline algorithms are statistically significant. On average PSC outperforms baselines B1 and B2[3] by 49% and 26%, respectively.

LFK-NMI is a quantitative measure that captures the overlap between the algorithmic *clusters* and the *classes* in the ground truth. It reports a single-number summary, but it does not provide any details about the resemblance between clusters and classes in terms of their numbers and size. For instance, if there is a huge class in the ground truth along with several small ones, an algorithm can achieve high LFK-NMI by assigning all the data points to a single cluster. To investigate the performance in greater detail, let us consider the confusion matrix containing the Jaccard coefficient between the set of tweets of every cluster in the solution and in the ground truth, respectively. Fig. 5.2.2 shows the confusion matrices for the three algorithms. The rows and columns in these matrices represent the clusters in the solution and classes in the ground truth, respectively. The number next to each row (resp., column) shows the number of tweets in each cluster (resp., class). These matrices are computed at an evaluation period in which all three algorithms display local maxima in LFK-NMI. Although this period does not represent the best quality for any of the algorithms, it has the advantage that the ground truth classes are the same for all three algorithms, which is crucial for performance comparison.

A good clustering solution will have a confusion matrix with a dark colored cell (high value of Jaccard Coefficient) in each row or column. The perfect clustering would show

[3] **Baseline B1**: This configuration is an implementation of the Online K-means clustering of simple tweets along with outlier handling as explained earlier. The only feature used in this algorithm is text content. The Term Frequency (TF) vector of each tweet is used to compute the content similarity between tweets and aggregate them. **Baseline B2**: This configuration is an implementation of the event detection system recently proposed by Aggarwal and Subbian, which is a tweet clustering algorithm based on a combination of content and network features. To the best of our knowledge, this approach represents the current state of the art in streaming clustering of tweets. It relies on the full knowledge of the follower network of all users present in the dataset. Such information provides a very significant advantage, but it also creates a practical challenge in that it is very time-consuming to obtain, making the algorithm infeasible in real-time, streaming scenarios. To compute tweet similarity, the original algorithm adopts TF-IDF, but we use TF on our implementation as it provides better performance on our dataset. This algorithm is also based on Online K-means and incorporates the same outlier handling procedure. To make use of this algorithm for comparison, we extracted in batch the follower network of all users present in our dataset.

only dark cells on the diagonal of a square confusion matrix. As Fig. 5.2.2 illustrates, PSC does a good job at capturing the actual clusters in the data; we observe greater confusion in the clusters generated by the two baseline algorithms. In particular, our method is able to recover 8 clusters whose overlap with the ground truth cluster is above 60%, while both the baseline methods identify at most 3 clusters faithfully resembling the ground truth. Although the performance of the clustering methods fluctuates over time, PSC is able to outperform the state of the art and discover memes in a streaming scenario with reasonable accuracy.

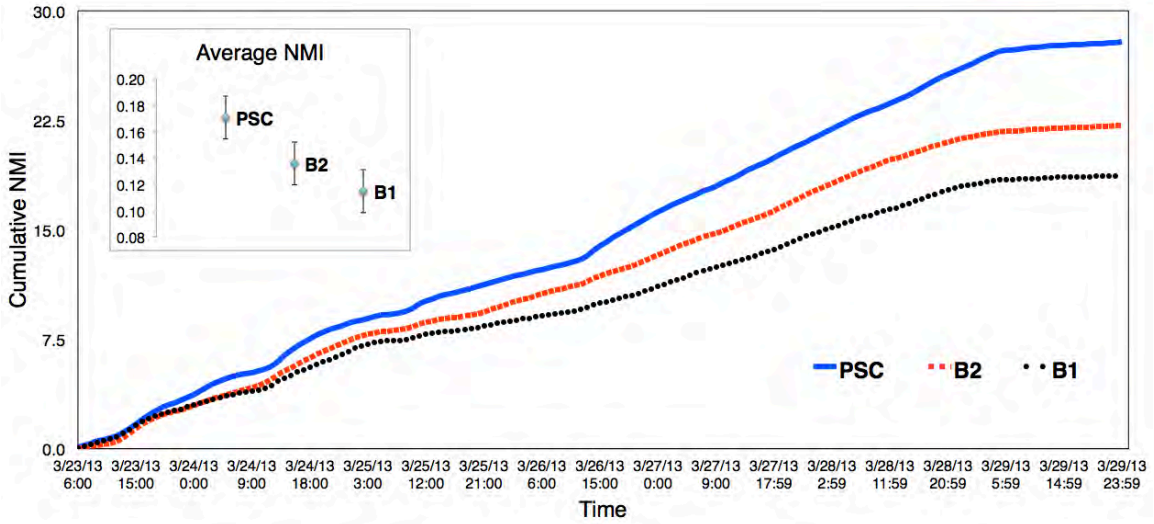



**Fig. 5.2.1**: Performance of different clustering algorithms as a function of the evaluation period. For each algorithm, the LFK-NMI values at each step are averaged across five runs. These values are then accumulated over the course of the experiment. The inset plots the time-averaged LFK-NMI, with error bars corresponding to $\pm 1$ standard error.

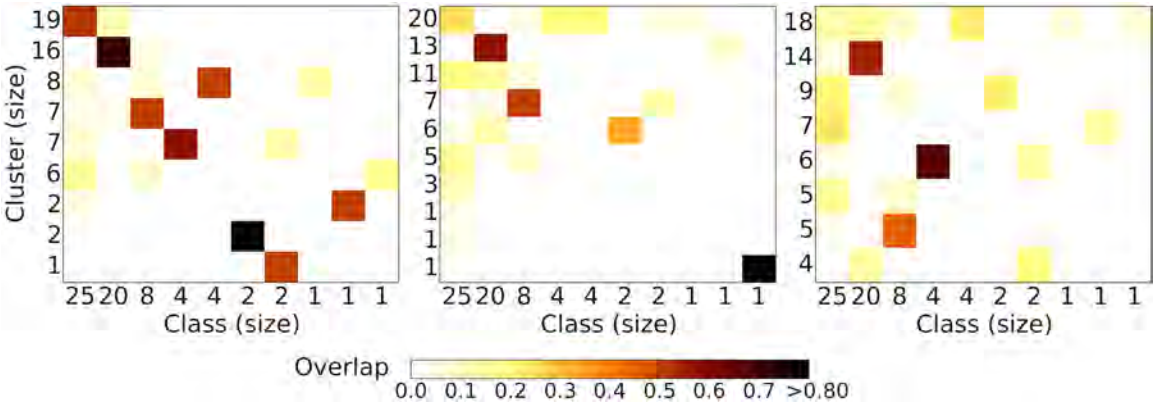



**Fig. 5.2.2**: Overlap (Jaccard coefficient) between ground truth classes and clusters detected by PSC (left), B2 (middle), and B1 (right).

### 5.3 Optimal modular structure for information diffusion in social networks

As the main goal of the DESPIC project was to produce a framework to identify coordinated efforts to spread (mis-) information, we felt it was necessary to study, on a theoretical ground, the general mechanisms that drive the spread of (mis)information and rumors. The IU team worked extensively to model diffusion dynamics in networks with realistic structure.

Following recent findings on the spread of behaviors on social networks,[4] our group investigated what is the role of the community structure in presence of social reinforcement. Social reinforcement provisions that each additional exposure to a piece of information sensibly increases the probability of its adoption. This makes diffusion phenomena in social networks behave differently from simple spreading, say e.g., epidemics. Epidemic spreading is hindered by the presence of communities or modular structure, since this helps confining the epidemics in the community of origin. We investigate whether this holds true for information spread with social reinforcement.

We exposed the two, somewhat antagonistic, effects determined by the modular structure of the social network: *enhancement of local spreading* and *hindrance of global spreading*. Strong communities facilitate social reinforcement and thereby enhance local spreading; weak community structure makes global spreading easier, because it provides more bridges among communities. We show that there exists an *optimal balance* between these two effects, where community structure counter intuitively *enhances* ---rather than hindering--- global diffusion of information.

#### 5.3.1 Methods

We use the linear threshold model ---which incorporates the simplest form of social reinforcement--- to systematically study how community structure affects global information diffusion. Consider a set of $\mathcal{N}$ nodes (agents) connected by $\mathcal{M}$ undirected edges. The state of an agent $i$ at time $t$ is described by a binary variable $s_i(t) = \{0,1\}$, where 1 represents the `active' state and 0 the `inactive' one. At time $t$=0 a fraction $p_0$ of randomly selected agents, or `seeds,' is initialized in the active state. At each time step, every agent's state is updated synchronously according to the following threshold rule:

$$s_j(t+1) = 1 \text{ if } \theta k_i < \sum_{j \in N(i)} s_j(t),$$
$$s_j(t+1) = 0 \text{ otherwise,}$$

where $\theta$ is the threshold parameter, $k_i$ is the degree of node $i$, and $N(i)$ is the set of $i$'s neighbors. Note that a node turning active cannot become inactive, and that the diffusion dynamics is deterministic. The system reaches a steady state when no further activations are possible.

---

[4] Centola, D. The spread of behavior in an online social network experiment. Science, 2010.

We create an ensemble of networks with modular structure using the block model approach, assigning half of the nodes to one community and the remainder to a second one. The interconnectivity within and between communities is tuned by means of a parameter $\mu$. Larger $\mu$ indicate strong inter-community connectivity, and smaller $\mu$ the vice-versa. We initiate the spreading from one community activating a fraction of nodes and resolve the diffusion equations using two approximation methods: (i) mean field (MF), and (ii) tree-like (TL) approximations. In the former case, the equation to compute the smaller stable solution for the fraction of active nodes $p_\infty^A$ in community A is written as:

$$p_\infty^A = p_0^A + (1 - p_0^A) \sum_{k=1}^{\infty} p(k) \times \sum_{m=\lceil \theta k \rceil}^{k} \binom{k}{m} (q^A)^m (1 - q^A)^{(k-m)}$$

where $p_0^A$ is the density of the seeds in the community A, and $q^A = (1-\mu)p_\infty^A + \mu p_\infty^B$ is the probability that neighbor of a node is active, which is the sum of: (i) the probability that the neighbor is in the same community $(1-\mu)$ and is active $(p_\infty^A)$; and, (ii) the probability that it is in the other B community $(\mu)$ and is active $(p_\infty^B)$.
It is straightforward to write a symmetric equation to compute $p_\infty^B$. Finally, $p_\infty = (p_\infty^A + p_\infty^B)/2$.

The more sophisticate TL approximation maps the underlying network into a tree of infinite depth and assumes that nodes at level *n* are only affected by those at level *n-1*. For the details of the solution computed for this approximation refer to [3].

### 5.3.2 Results

We addressed the issue of how communities affect information diffusion. As $\mu$ decreases nodes in A have increasingly more neighbors in A. Thus, the number of seed nodes to which nodes in A are exposed also increases because the seeds exist only in A. In other words, strong communities *enhance* local spreading. By contrast, the spreading in community B is triggered entirely by the nodes in A. Therefore, larger $\mu$ (smaller modularity) helps the spreading of the contagion to community B. The fact that large modularity (smaller $\mu$) facilitates the spreading in the originating community, but small modularity (larger $\mu$) helps inter-community spreading, raises the following question: is there an optimal modularity that facilitates both intra- and inter-community spreading?

Our work suggests a positive answer to the question raised above. Fig. 5.3.2.1 demonstrates that there is indeed a range of values of $\mu$ that enables both. In the blue range (``local''), strong cohesion allows intra-community spreading in the originating community A; in the red range (``global''), weak modular structure allows inter-community spreading from A to B. The interval in which blue and red overlap (purple, ``optimal'') provides the right amount of modularity to enable global diffusion.

Fig. 5.3.2.2 summarizes our results, derived analytically by MF and TL approximations, and by numerical simulations. We compute the mean of $p_\infty$ across 1,000 runs of the model, each assuming a different realization of the network and of the seed nodes. We fix the threshold ($\theta = 0.4$) throughout all simulations. Fig. 5.3.2.2(a) shows the phase diagram with three phases: no diffusion (white), local diffusion (blue), and global diffusion (red).

These results, published in the prestigious *Physical Review Letters* [3], have been selected as "Editors' Suggestion", an honor awarded only to one article per issue.

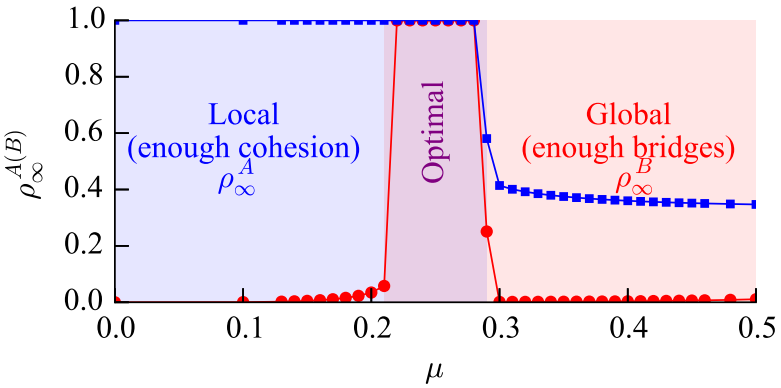



**Fig. 5.3.2.1**: The tradeoff between intra- and inter-community spreading. Stronger communities (small $\mu$) facilitate spreading within the originating community (local) while weak communities (large $\mu$) provide bridges that allow spreading between communities (global). There is a range of $\mu$ values that allow both (optimal). The blue squares represents $p_\infty^A$, the final density of active nodes in the community A, and the red circles represents $p_\infty^B$. The parameters for the simulation are: $p_0 = 0.17$, $\theta = 0.4$, N = 131,056 and $z$ = 20.

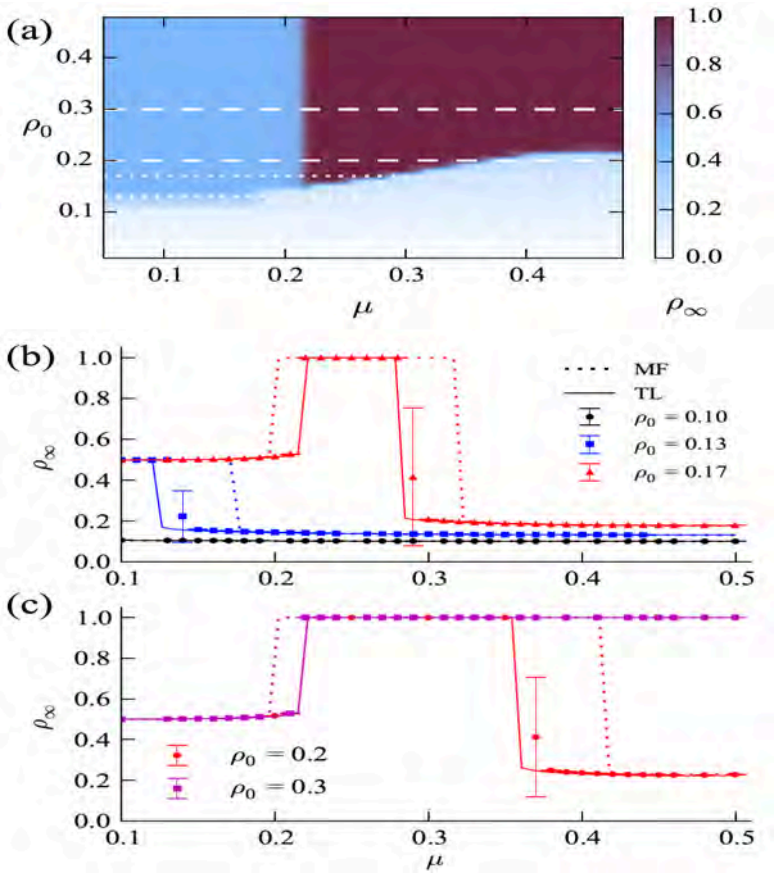



**Fig. 5.3.2.2**: (a) the phase diagram of threshold model in the presence of community structures with $N$ = 131,056 and $z$=20, and $\theta = 0.4$. There are three phases: no diffusion (white), local diffusion that saturates the community A (blue), and global diffusion (red). The dotted and dashed lines indicate the values of $p_0$ shown in (b) and (c). (b) the cross-sections of the phase diagram (dotted lines in (a)). TL (solid lines) shows excellent agreements with the simulation while MF (dotted lines) overestimate the possibility of global diffusion. (c) the cross-sections represented in dashed lines in (a).

## 5.4 Evolution of online user behavior, roles and influence

During the three years of the DESPIC project the IU Team started focusing on the role played by single users in spreading information and how they influence each other. The ultimate goal is that of understanding the mechanisms behind successful persuasion campaigns. We conducted several case study to learn what features are revealing of the influence process and of the users' role.

### 5.4.1 Evolution of online user behavior

We conducted a case study focused on the protests that occurred in Turkey in 2013 (the so called Gezi Park and Taksim Square upheavals) and the relative social media discussions.

We collected a dataset of tweets isolating 32 hashtags of general interest for the protest, including 10 among those adopted by the protesters and 10 among those adopted by government supporters. We expanded this list by extracting the 100 hashtags that occurred more frequently with those in the seed list. We finally gathered (from the Twitter Gardenhose – a 10% sample of the entire social media stream) all the tweets containing at least one hashtag in the extended list. This procedure yielded a dataset of 2.3 million tweets produced during the 25 days of the protest between May and June 2013.

The details of our analysis are reported in [6]: our work was presented at the prestigious ACM Web Science 2014 Conference and was praised with the conference Best Paper Award.

The most relevant findings of our study are summarized here: (i) we presented a method to extract topically focused conversation on the Gezi Park protest; (ii) we explored the spatio-temporal characteristics of such conversation, in particular we studied where tweets originated and where they were consumed. This allowed us identifying clusters of cities that are mostly consistent with the country geopolitics; (iii) we analyzed the emerging characteristics of the users involved in this conversation, including their roles and their influence, and showed how these evolved as the events of the protest unfolded, highlighting a redistribution of influence that made the discussion more democratic and less centered around some key actors; (iv) we showed how online user behavior is affected by external events and how users respond to political leaders' speeches with the emergence of a spontaneous synchronization process.

We first explored the spatial dimension of the conversation, focusing on the discussion inside the Turkish borders. In Fig. 5.4.1.1 we show the trend similarity matrix computed among the sets of trending hashtags and phrases occurring in each of the 12 cities where Twitter trends are monitored.
The clusters found on the base of the similarity matrix matches well-known fact relative to the Turkish geopolitical profile. Eskisehir, Kayseri and Gaziantep (in the red cluster) are all central Anatolian cities where the president's party (AKP) has a stronghold (though the CHP opposition party edged out the AKP in the March 2014 mayoral race); they are more culturally conservative and homogeneous. Izmir, Istanbul, Bursa, Ankara, and Adana (green cluster) are the largest cities in Turkey with diverse populations. Finally, Antalya and Mersin (blue cluster) are seacoast cities that are known for supporting either one of the main opposition parties (CHP or MHP).

We then explored the temporal dimension of the Gezi Park discussion. The aim was to determine whether the activity on social media mirrored on-the-ground events, and whether bursts of online attention coincided with real-world protest actions. We analyzed the time series of the volume of tweets, retweets and replies occurring during the 27-day-long observation window, as reported in Fig. 5.4.1.2 (top panel). The discussion was driven by bursts of attention that largely corresponded to major on-the-ground events, similar to what has been observed during other social protests. The numbers of tweets and retweets are comparable throughout the entire duration of the

conversation, suggesting a balance between content production (writing novel posts) and consumption (reading and rebroadcasting posts via retweets). In the middle panel of Fig. 5.4.1.2 we report the number of users involved in the conversation at a given time, and the cumulative number of distinct users (dashed red line). Similarly, in the bottom panel, we show the total number of hashtags related to Gezi Park, and the cumulative number of distinct hashtags. Approximately 60% of all users observed during the entire discussion joined in the very first few days, whereas additional hashtags emerged at a more regular pace throughout a longer period. This suggests that the conversation acquired traction immediately, and exploded when the first on-the-ground events and police action occurred.

Our second experiment aims at investigating what roles users played in the Gezi Park conversation and how they exercised their influence on others. We also seek to understand whether such roles changed over time, and, if so, to what extent such transformation reshaped the conversation. Fig. 5.4.1.3(a) shows the distribution of social ties reporting the two modalities of user connectivity, namely followers (incoming) and followees (outgoing) relations. The dark cells along the diagonal indicate that most users have a balanced ratio of ingoing and outgoing ties. Users below the diagonal follow more than they are followed. Note that most users are allowed to follow at most 1000 people. Finally, above the diagonal, we observe users with many followers. Note the presence of extremely popular users with hundreds of thousands or even millions of followers. The number of followers has a broad distribution and seems largely independent of the number of followees. The presence of highly followed users in this conversation raises the question of whether their content is highly influential. We determined user roles as a function of their social connectivity and interactions. Fig. 5.4.1.3(b) gives an aggregated picture of the distribution of user roles during the Gezi Park conversation. The y-axis shows the ratio between number of followees and followers of a given user; the x-axis shows the ratio between the numbers of retweets produced by a user and the number of times other users retweet that user. In other words, the vertical dimension represents social connectivity, whereas the horizontal dimension accounts for information diffusion. We can draw a vertical line to separate influential users on the left (those whose content is most often retweeted by others) and information consumers on the right (those who mostly retweet other people's content). Influential users can be further divided in two classes: those with more followers than followees (bottom-left) and those with fewer followers (top-left), which we call hidden influentials. Similarly, information consumers can be divided in two groups--rebroadcasters with a large audience (bottom-right), and common users (top-right). Fig. 5.4.1.3(b) shows a static picture of aggregated data over the 27-day observation period.

To study how roles evolve as events unfold, we carried out a longitudinal analysis whose results are shown in Fig. 5.4.1.4. This figure shows the average displacement of each role class, and the number of individuals in each class (circles), for each day. The displacement is computed in the role space (that is, the space defined by the two dimensions of Fig. 5.4.1.3(b)). Larger displacements suggest that individuals in a class, on average, are moving toward other roles.

Various insights emerge from Fig. 5.4.1.4: first, we observed that the classes of information producers (influentials and hidden influentials) are relatively stable over time; together they include more than 50% of users every day, suggesting that a consistent number of individuals had large audiences, and the content they produced was heavily rebroadcasted (by information consumers as well as other influentials). On the other hand, information consumers show strong fluctuation: starting from an initial configuration with stable roles (May 29--31), common users and rebroadcasters subsequently exhibit large aggregate displacements in the role space (June 1--4). We also note a redistribution of the users in each role: at the beginning of the protest a large fraction represents common users and rebroadcasters, while, as time passed and events unfolded, these two classes shrank. This suggests that common users and rebroadcasters acquired visibility and influence over time: some fraction of these users moved from the role of information consumers to that of influentials, such that their content was consumed and rebroadcasted by others. In other words, the discussion became *more democratic* over time, in that the control of information production was redistributed to a larger population, and individuals acquired influence as the protests unfolded.

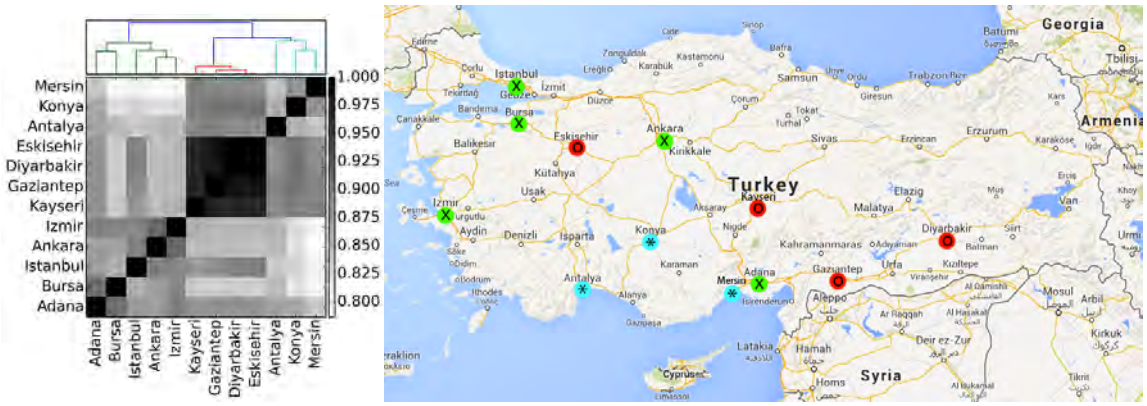



**Fig. 5.4.1.1**: (left) Trend similarity matrix for 12 cities in Turkey. From the dendrogram on top we can isolate three distinct clusters. (right) Location of the cities with trend information, labeled by the three clusters induced by trend similarity.

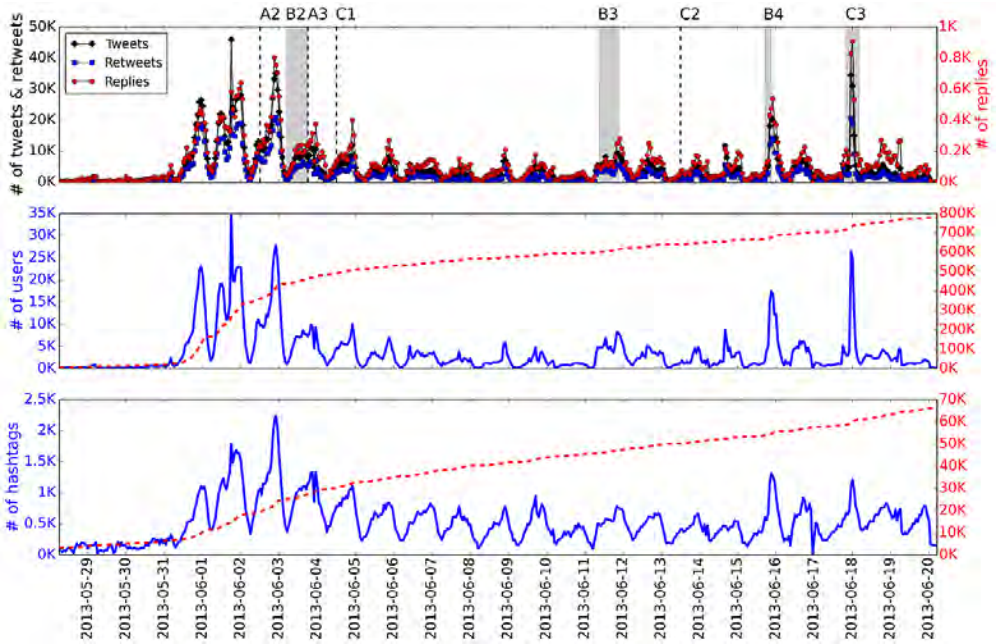



**Fig 5.4.1.2**: Hourly volume of tweets, retweets and replies between May 30th and June 20th, 2013 (top). The timeline is annotated with events from Table 1 of ref. [6]. User (center) and hashtag (bottom) hourly and cumulative volume of tweets over time.

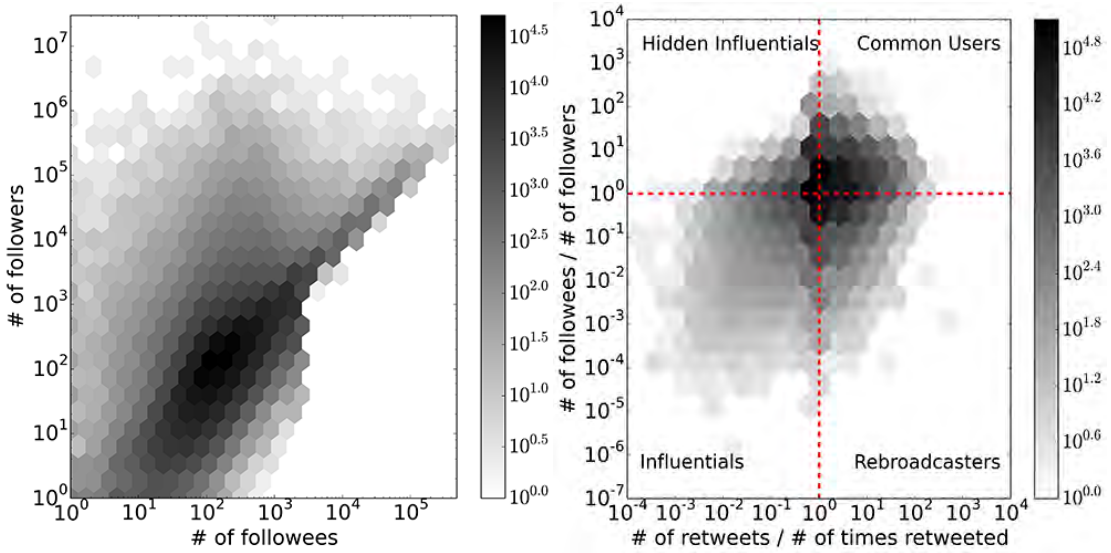



**Fig. 5.4.1.3**: (left) Distribution of friends and followers of users involved in the Gezi Park conversation; (right) Distribution of user roles as function of social ties and interactions.

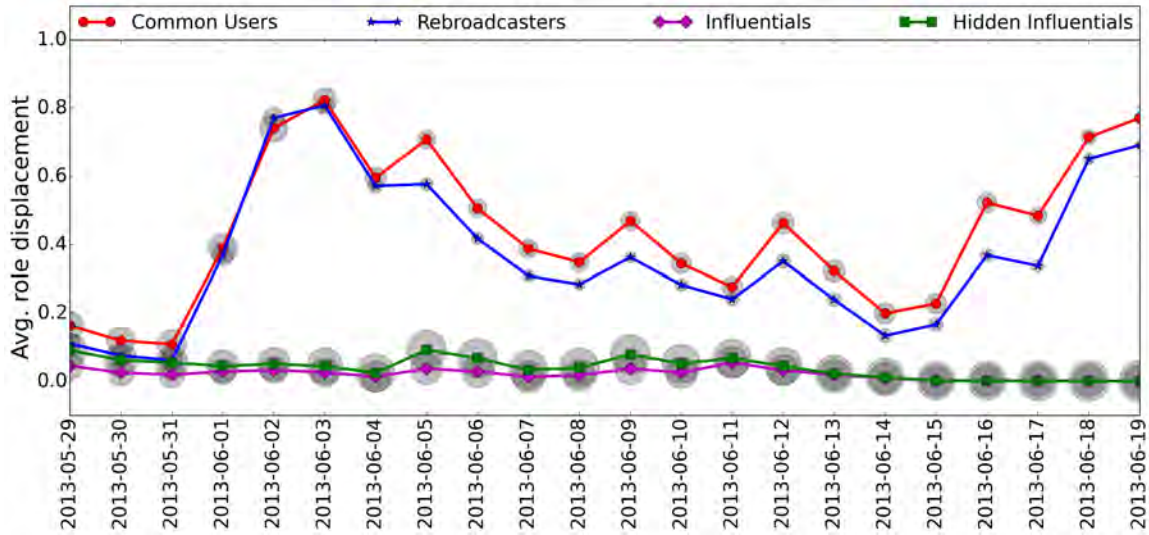



**Fig. 5.4.1.4**: Average displacement of roles over time for the four different classes of roles. The size of the circles represents the number of individuals in each role.

### 5.4.2 Grassroot meme formation and evolution analysis

The IU Team has investigated how grassroot memes form and evolve over-time. Studying a specific meme about the social movement known as "Occupy Wall Street" since its inception and for the duration of one year, we sought to understand: (1) how the geographic patterns of this communication network differ from those of stable political communication [9]; (2) how Twitter users taking part to the protest took up different online behaviors, language usage and activity modes, and (3) how online protest groups were formed and how they evolved.

We found that, compared to a network of stable domestic political communication, the Occupy Wall Street grassroot meme exhibits higher levels of locality and a hub and spoke structure, in which the majority of non-local attention is allocated to high-profile locations such as New York, California, and Washington D.C. This signal might be of extreme importance to determine a signature of a genuine grassroot movement, encoding the importance of the geographic dimension in the characterization of memes' nature.

Moreover, we observed that information flowing across state boundaries is more likely to contain framing language and references to the media, while communication among individuals in the same state is more likely to reference protest action and specific places and times, as reported by the Figure 5.4.2.1 below. This uniquely characterizes the signature of a grassroot meme yielding to a differentiation of the content diffused and language adopted on the geographical network according to different scopes.

| Interstate | | Intrastate | |
|---|---|---|---|
| **Token** | **Ratio** | **Token** | **Ratio** |
| wall | .590 | city | 2.254 |
| nyc | .600 | tonight | 1.737 |
| street | .699 | march | 1.669 |
| news | .718 | join | 1.494 |
| 99% | .756 | solidarity | 1.387 |
| bank | .763 | day | 1.354 |
| don't | .782 | square | 1.333 |
| media | .837 | please | 1.243 |
| peaceful | .845 | park | 1.220 |
| nypd | .847 | now | 1.179 |

'Ratio', defined as $\frac{P(Token|Intrastate)}{P(Token|Interstate)}$, is small when a token is more common in intrastate traffic and large when a token is more common in interstate traffic. Terms relating to rallying supporters are more predominant in intrastate communication, while interstate traffic tends to favor terms such as protest slogans and references to the media.

**Fig. 5.4.2.1**: Geographical differences in language usage in a grassroot meme (#ows).

We finally examined the temporal evolution of digital communication activity relating to the protest observing the changes in users' engagement, interests, and social connectivity. The results of this analysis indicated that, on Twitter, the movement tended to elicit participation from a set of highly interconnected users with pre-existing interests in domestic politics and foreign social movements (see Figure 5.4.2.2). These users, while highly vocal in the months immediately following the birth of the movement, appeared to have lost interest in protest over the remainder of the study period. Our findings related to topical group formation and evolutions in a grassroot meme are instrumental to understanding what components play an important role in the characterization of natural campaigns vs. an artificial one.

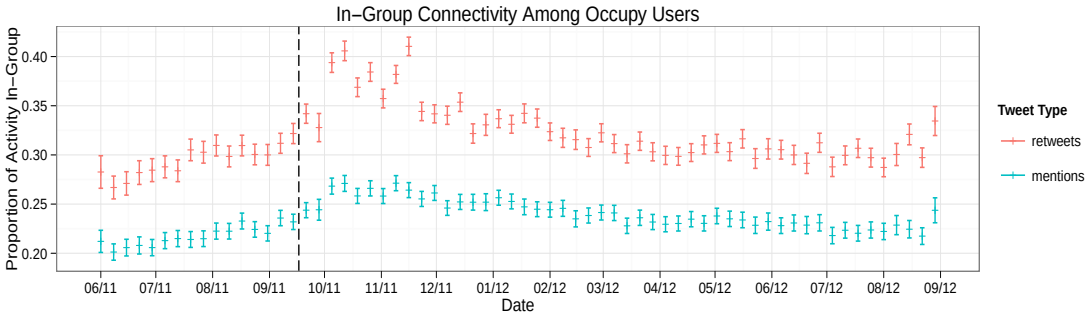



**Fig. 5.4.2.2**: Group formation and connectivity evolution for the users involved in the #ows conversation.

We further investigated the birth and growth of natural trends, exploiting the above-mentioned dataset built for the artificial vs. natural trend classification problem, trying to generalize our findings to encompass all observed trends. In particular we focused on understanding the role played by geography in shaping the communication patterns describing these naturally trending memes.

Focusing once again on United States trends, looking for universal dynamics to describe the spread of grassroot trending memes, we found that two very different classes of dynamics exist: (1) trend-setting, and (2) trend-following ones. I.e., certain cities that correspond to the largest traffic hubs of the country play the role of trend-setters, producing a significantly larger fraction of trends that will later be followed by other states; on the other hand, most of the other cities act as trend-followers, mostly receiving the trends from the trend-setters and rebroadcasting them rather than producing new trends.

These two clearly separate dynamics are shown in the Figure 5.4.2.3 below.

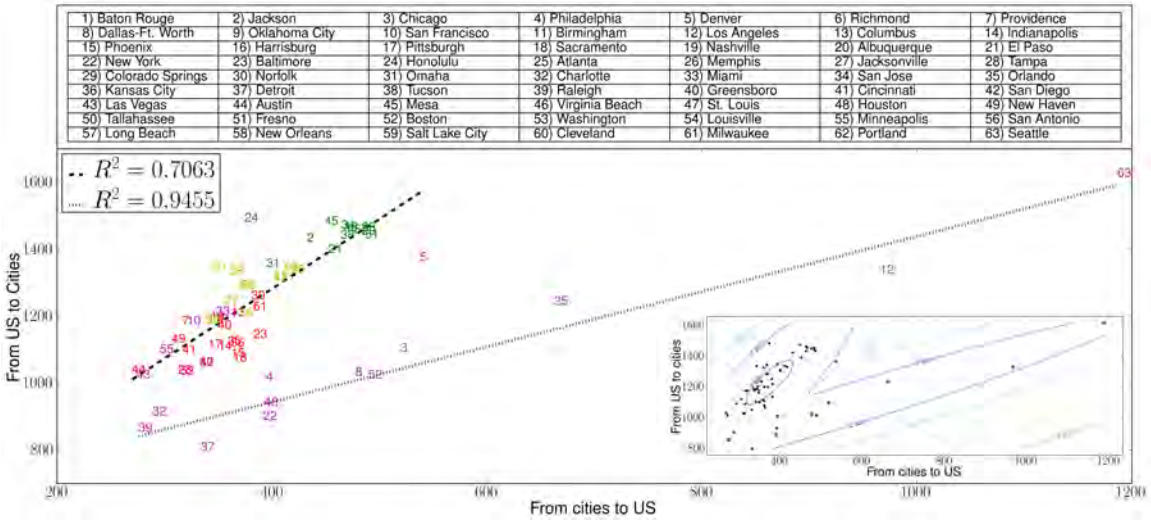



**Fig 5.4.2.3.** Trendsetters vs. trend-followers: the inset shows a Gaussian Mixture Model showing two different trendsetting dynamics; the contours represent the std. dev. of each Gaussian distribution. The main plot shows their linear regressions.

Although not originally included in our proposal, we decided to explore the possibility to classify users behavior to increase our arsenal of tools to detect the variety of (non-) malicious behaviors underlying the diffusion of information in online social media. The motivating hypothesis beyond this choice is that knowing who are the users participating a conversations (knowing, for example, what are their historical interests, political affiliations, friends, posting habits) provides information about the nature of the conversation that cannot be found in the topological and textual features of the same users when observed in isolation.

While collecting users' history is within our capabilities (modulo technical impediments related to data size and their streaming nature), what is direly needed is a framework that can summarize users' behavior in a number of meaningful classes.

As a case study we attempted the classification of users according to their behavior in establishing friendship relationship on an online social media [21, 30]. We used proprietary data provided by colleagues at Yahoo! Research in Barcelona that describes the full and detailed history of users' link creation and message sending/repost in Yahoo! Meme (a platform similar to Twitter, now discontinued, used mostly for sharing images).

We identified a number of simultaneous potential strategies a generic user may adopt to create new links (e.g., to follow the friend of a friend, or follow the user originating messages that the link creator often reposts – and therefore finds interesting). Each user is then represented with a vector of probabilities that describes how often the user draws her linking choices from the corresponding strategy. These probabilities are then estimated assuming that the observed history of the entire network maximizes the likelihood of what is observed among all possible values of the individual strategy probabilities. Each user is finally described by the relative frequency with which she uses the basis strategies. Users are then clustered according to their probability vector in groups with similar behavior.

From this study it emerges that a successful classification is possible. Five different well-distinct classes emerge that greatly differ not only in the dimensions we chose a-priori to describe them, but in several others, including their permanence on the platform, the frequency of their posts, how much they are followed and reposted, and more. The differences between the different groups across a number of these dimensions are illustrated in Figure 5.4.2.4. Notice for example the group we named *information seekers* (pink). Their linking behavior is driven by the trial to reach directly the source of information of interests rather then expanding their friendship circle and aligning their interest on it. Although they are a minority (3%) they produce more information, and even more importantly they act as spreaders of the information they collect widely across the network rather then just in the restricted circle of their friends.

The analytical framework we developed for this study is relatively simple and general enough to be extended to a number of behaviors beyond how online follower/followee relations are established. Of course its reliability rests on the careful choices of the probability vectors that describe the users and what they represent. Those will need to be chosen by careful consideration depending on the specific user feature being analyzed.

The results of this study were presented at the prestigious KDD 2013 conference, held in Chicago in September 2013.

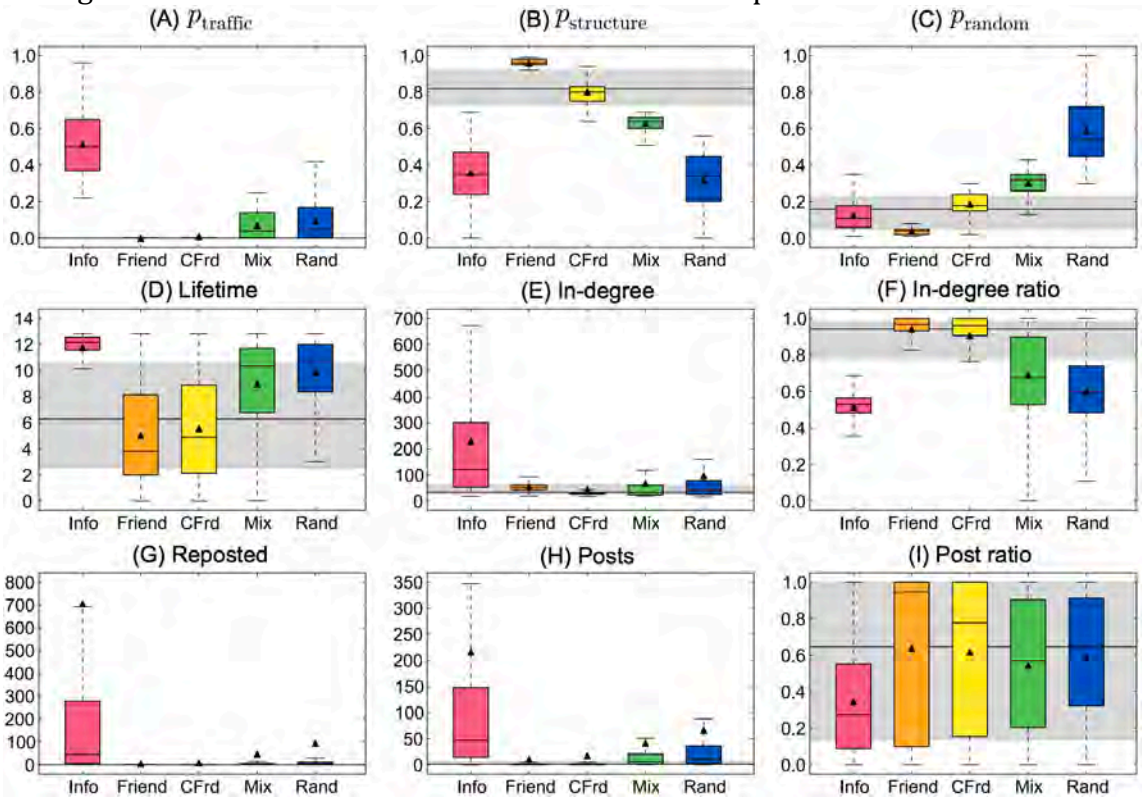



**Fig. 5.4.2.4**: Features describing different classes of users. Each box shows data within lower and upper quartile; whiskers represent 99th percentile; the triangle and the line in a box represent median and mean, respectively.

## 5.5 Social bot detection

Persuasion campaigns are often accompanied by an intense usage of social bots. Another milestone achieved by the IU Team is the design, development and delivery of an automatic framework for detection and classification of human and synthetic activity on social media.

### 5.5.1 Bot or not?

Our team worked on a machine-learning framework to extract features that characterize human-like and bot-like behavior, exploiting various dimensions of account activities. Our system exploits over one thousands features derived from user meta-data, content and sentiment produced and consumed, timing information, network structure and information diffusion patterns. For each feature the system builds a set of descriptive statistics that include mean, moments, and the entropy of the distribution.

Features are divided in the following classes: (i) *Network features* capture various dimensions of information diffusion patterns. We build networks based on *retweets*,

*mentions*, and *hashtag co-occurrence*, and extract their statistical features. Examples include degree distribution, clustering coefficient, and centrality measures. (ii) *User features* are based on Twitter meta-data related to an account, including language, geographic locations, and account creation time. (iii) *Friend features* include descriptive statistics relative to an account's social contacts (followees), such as the median, moments, and entropy of the distributions of their number of followers, followees, posts, and so on. (iv) *Timing features* capture temporal patterns of content generation (tweets) and consumption (retweets); examples include the signal similarity to a Poisson process, the average time between two consecutive posts, and such. (v) *Content features*, are based on linguistic cues computed through natural language processing, especially part-of-speech tagging; examples include the frequency of verbs, nouns, and adverbs in the phrases produced by the account. (vi) *Sentiment features* are built using general-purpose and Twitter-specific sentiment analysis algorithms, including happiness, arousal-dominance-valence, and emotion scores.

When our framework analyzes an account, it creates a "profile" extracting all these features, and then it runs the profile through a set of classifiers that include decision trees, ensemble methods (random forest), boosting methods (AdaBoost) and linear models (Logistic Regression). The models can classify the profile using all features combined or using disaggregated feature classes, considering only one set of features from the same class at the time.

To classify an account as either social bot or human, the models must be trained with instances of both classes. Finding and labeling many examples of bots is challenging. As a proof of concept, we used a list of social bots compiled by Lee et al.[5] We used the Twitter Search API to collect up to 200 of their most recent tweets and up to 100 of the most recent tweets mentioning them. This procedure yielded a dataset of 15 thousand manually verified social bot accounts and over 2.6 million tweets. Lee's list also contains legitimate (human) accounts. The same procedure resulted in a dataset of counterexamples with 16 thousand people and over 3 million tweets. We used this dataset to train the social bot detection model and benchmark its performance.

"*Bot or Not?*" achieves very promising detection performance, with a ROC-AUC score of 95% (see Fig. 5.5.1.1). Some feature classes, like the user meta-data, appear more revealing and they can be easily explained (see Fig. 5.5.1.2). Note that such performance evaluation is based on Lee's dataset from 2011; we are already aware of more recent social bots that cannot be reliably detected. Bots are continuously changing and evolving. Further work is needed to identify newer annotated instances of social bots at scale. The DARPA SMISC Synthetic Account Detection Challenge will represent an optimal testbed for our framework to measure it's detection performance with a third-party benchmark.

To make the detection system broadly accessible, we developed a Web-based application that interfaces with the Twitter API and retrieves the most recent activity of any account,

[5] Lee, K. et al. Seven months with the devils: a long-term study of content polluters on Twitter. ICWSM 2011

to make a determination of whether that account exhibits bot-like or human-like behavior. The Web interface, depicted in Fig. 5.5.1.3, allows one to inspect any active Twitter account. Data about that account and its contacts are collected and processed in real time. The classifier trained on all feature classes provides a likelihood score that the account is a social bot. The system also presents disaggregate scores according to models trained on each feature class independently. Often, an account may be classified as a social bot according to some feature classes, but not according to others. This is due to the large heterogeneity of features exhibited by people ---some may have bot-like features, for example their meta-data or friend information. In addition to the classification results, "Bot or Not?" provides a variety of visualizations that capture some insights about the features exploited by the system. Examples are displayed in Fig. 5.5.1.3. We invite the reader to explore these interactive visualizations directly at http://truthy.indiana.edu/botornot

In summary, our research line on social bot detection led us to define and develop a new machine-learning framework to classify Twitter accounts as human-like or bot-like according to features describing their behavior, network characteristics, content and sentiment, and temporal patters, along with user meta-data. We also delivered as a proof-of-concept "Bot or Not?," a freely-accessible Web platform for social bot detection on Twitter. Our results are described in details in ref. [5] and are in revision under the prestigious journal *Communications of the ACM*.

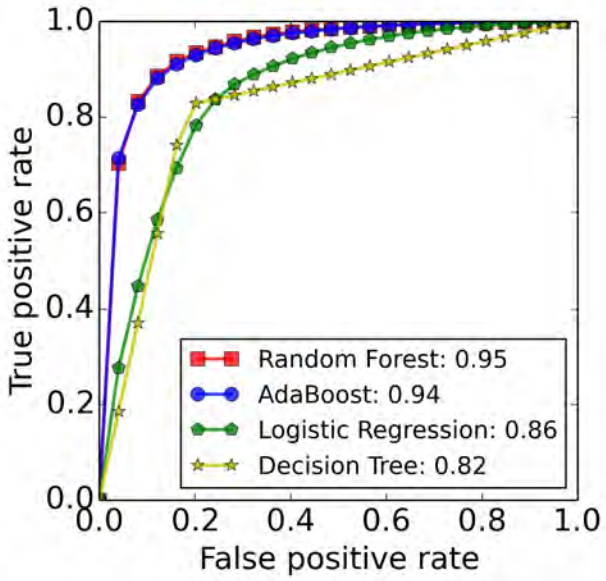



**Fig. 5.5.1.1**: Classification performance of "*Bot or Not?*" for four different classifiers. The classification accuracy is computed by 10-fold cross validation and measured by the area under the receiver operating characteristic curve (AUROC). The best score, obtained by Random Forest, is 95%.

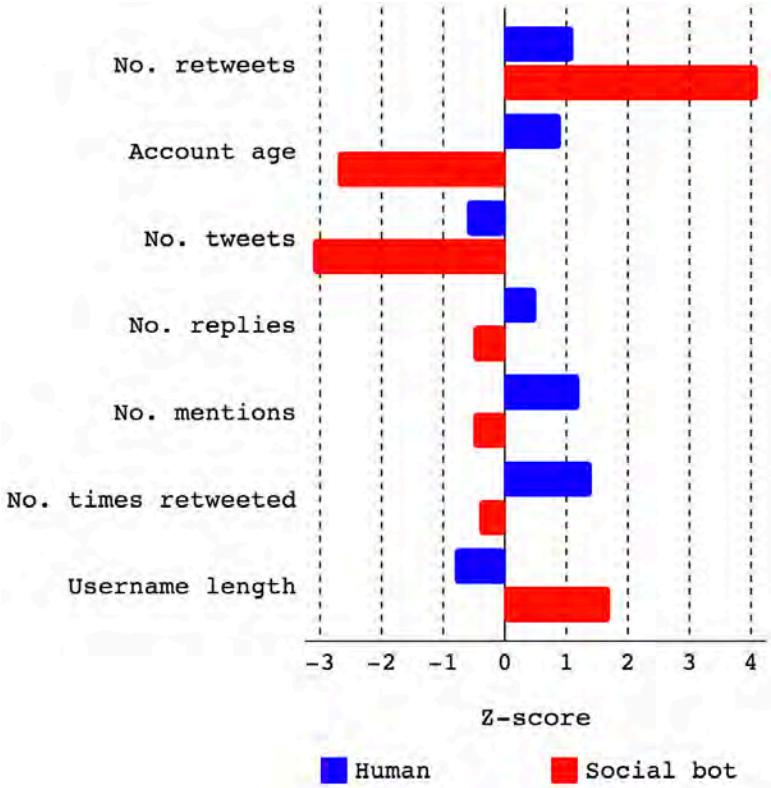



**Fig. 5.5.1.2**: Subset of user features that best discriminate social bots from humans. Bots retweet more than humans and have longer user names, while they produce fewer tweets, replies and mentions, and they are retweeted less than humans. Bot accounts also tend to be more recent.

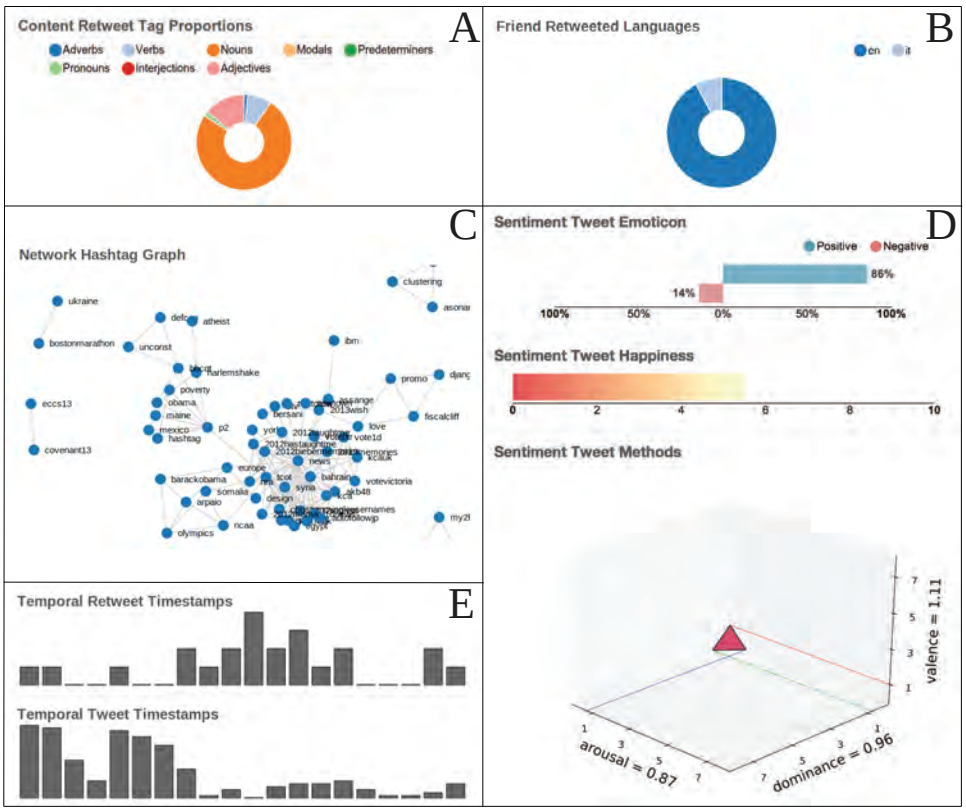



**Fig. 5.5.1.3**: Visualizations provided by "*Bot or Not?.*" (A) Part-of-speech tag proportions. (B) Language distribution of contacts. (C) Network of co-occurring hashtags. (D) Emotion, happiness and arousal-dominance-valence sentiment scores. (E) Temporal patterns of content consumption and production

### 5.5.2 SMISC Bot Detection Challenge

The main goal of the DARPA Bot Detection Challenge was to evaluate our capability to detect social media persuasion campaigns that are run by autonomous software, a.k.a. social bots. To this end, the Data Working Group of the DARPA SMISC program designed a realistic task that required participants to identify social bot accounts targeting anti-vaccine activists on Twitter. A "replay" of one month of data from the Twitter stream was provided to all challenge participants. The Indiana University team (including collaborators at the University of Michigan) designed a system to track, store and process the streaming data in real-time, while creating and updating the profiles of the accounts involved in the conversation, along with their corresponding features.

The IU approach consists of three steps:

1. Extraction of user-based features;
2. Filtering the search space based on various heuristics; and
3. Visualization and interactive data exploration for domain expert inspection.

### *Feature extraction*

The IU system builds a dynamic profile for each user participating in the conversation, for rapid data access, analysis, and classification. The system also generates feature vectors describing user profiles, updated every 6 hours, for classification purposes. It employes a subset of features developed for the 'Bot Or Not' framework (truthy.indiana.edu/botornot). The features can be summarized in five classes: user metadata, content, sentiment, network, and temporal features, as reported in Table 5.5.2.1. These features were carefully selected to reflect hand-crafted rules designed to identify suspicious activity. Examples of such rules include: (i) low entropy of topics of interest of the account, to identify thematically-focused users; (ii) anomalous levels of retweets or mentions, to capture users attempting to attract attention; (iii) anomalous connectivity patterns, to detect suspicious cliques; (iv) coordinated attempts to address specific human users, to identify orchestrated targeting; (v) suspicious growth-rate in followers, following, or content production levels; (vi) suspicious temporal patterns, as opposed of natural human circadian activity; (vii) high-volume of near-duplicate content; (viii) high-degree of sentiment polarization; and (ix) interactions focused on users in the target population, as opposed to external users.

As the stream of data was "replayed," the IU system periodically re-computed the user feature vectors. The pairwise cosine similarity between the feature vectors highlights the most similar pairs of users. Once the IU team started to identify bots in the conversation, matching the users most similar to the detected bots allowed for timely detection of new bots. In Fig. 5.5.2.1 we show the distribution of the pairwise cosine similarity between pairs of feature vectors characterizing bots, as opposed to bot-human pairs. The similarity between bots tends to be higher than between bots and humans. The bot-bot similarity exhibits a bimodal distribution that reflects the presence of two types of bots designed by two red teams: bots designed by same team are more similar to each other.

### *Heuristics*

In the earlier stage of the competition, the IU team developed various heuristic techniques to narrow the search space. Specifically, three strategies worked well: (i) analysis of the hashtag co-occurrence network; (ii) duplicate-image search; and (iii) dynamic tracking of network growth.

### Hashtag co-occurrence network

Starting from a provided list of vaccine-related hashtags, the IU team collected all tweets appearing in the competition stream that contained at least one of those hashtags. The system constructed a hashtag co-occurrence network, where each node represents a unique hashtag and edges between two nodes are weighted by the number of times these two hashtag are observed together in a tweet (see Fig. 5.5.2.2).

Using the hashtag co-occurrence networks, the IU team was able to identify other campaign-related hashtags to enrich the list of competition-relevant keywords. These were later used to separate users into categories of pro- and anti-vaccine. The proportion of tweets users posted containing any of these hashtags resulted in a strongly predictive feature.

### *Image search*

A common approach to create realistic bot profiles is to impersonate other users by cloning information such as descriptions, names, and profile pictures. The IU team built an algorithm to detect duplicate user pictures using an online image search service. Seven out of 39 bots were detected using this heuristics. These bots used images from the Wikipedia domain as their profile pictures.

### *Network growth*

In the competition dataset, a friendship network snapshot was provided every week. The IU team studied the topological changes of these temporal networks and identified users that created suspicious levels of connections with anti-vaccine activists.

### *Visualization*

Information visualization is a crucial part of the IU team's decision system. Expert knowledge is still required to conclude that a particular user is a social bot while limiting the number of false positives. The IU team developed a web application similar to the Twitter platform to create and populate user profile information and timelines in real time (see Fig. 5.5.2.3). This interface includes charts to monitor temporal changes in user metadata, such as the number of followers, friends, and posts.

### *Learning Framework*

The IU bot detection system relies on a strong machine-learning component, including several classifiers trained using previous bot datasets. We leveraged a training set from the Infolab at Texas A&M University, which consists of 15k verified social bot accounts and 16k legitimate users. The best features, a subset of those used in the 'Bot Or Not' system, are reported in Table 5.5.2.1. These features yielded a cross-validation classification accuracy of AUROC = 93% (Area Under Receiver Operating Characteristic curve). When applied to the challenge data, the classifiers identified several bots with high confidence, but unfortunately the majority of these were not challenge-related bots; they were not targeting anti-vaccine activists.

The lack of information about the actual number of bots in the challenge was one of the major challenges for the IU team. The target user set comprised more than 7000 users observed within the first two weeks of competition, only 39 of which (0.56%) were labeled as actual bots. We identified all of the bots, yielding no misses (zero false negative rate). However, we expected a higher number of bots. This brought the IU team to speculate, during the early weeks of the challenge, that a third class of bots remained undetected. We tried to balance exploration and exploitation seeking to reveal different classes of bots, yielding seven false hits (a false positive rate of 0.1%). A posteriori, a more conservative strategy would have produced fewer false alarms and better accuracy.

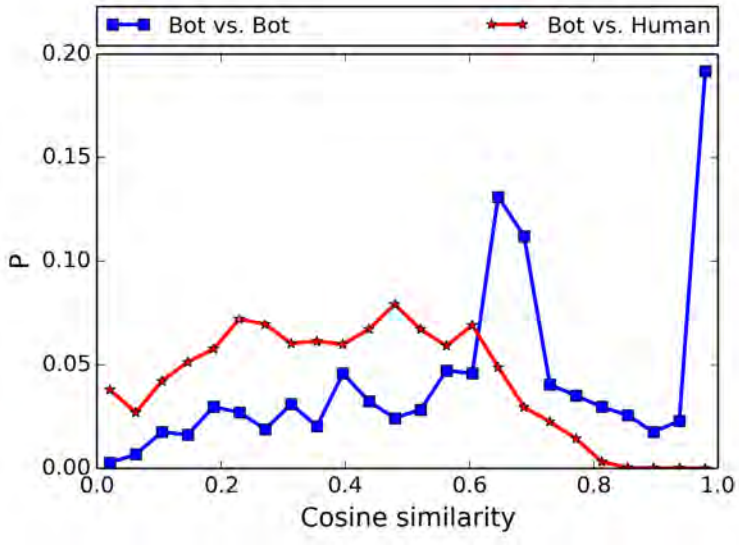



**Fig. 5.5.2.1**: Distribution of cosine similarity between pairs of accounts.

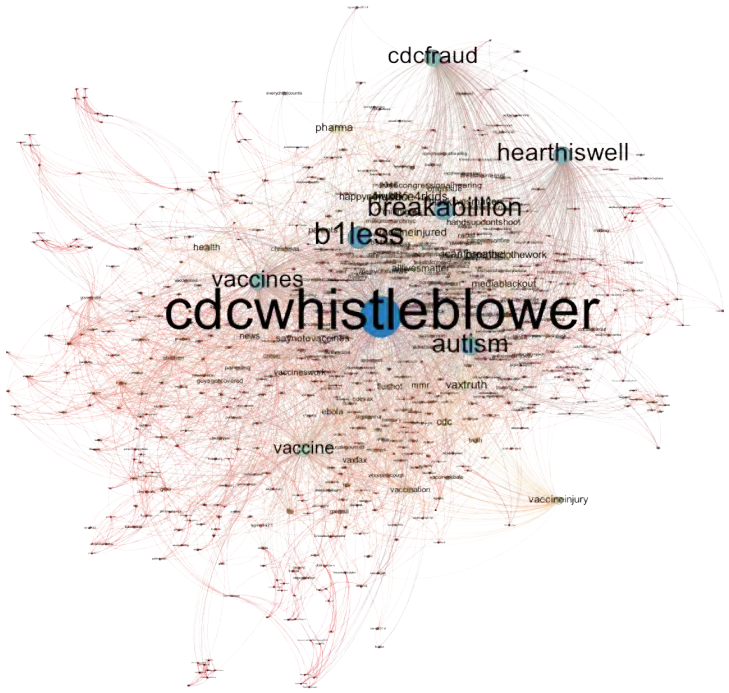



**Fig. 5.5.2.2**: Hashtag co-occurrence networks.

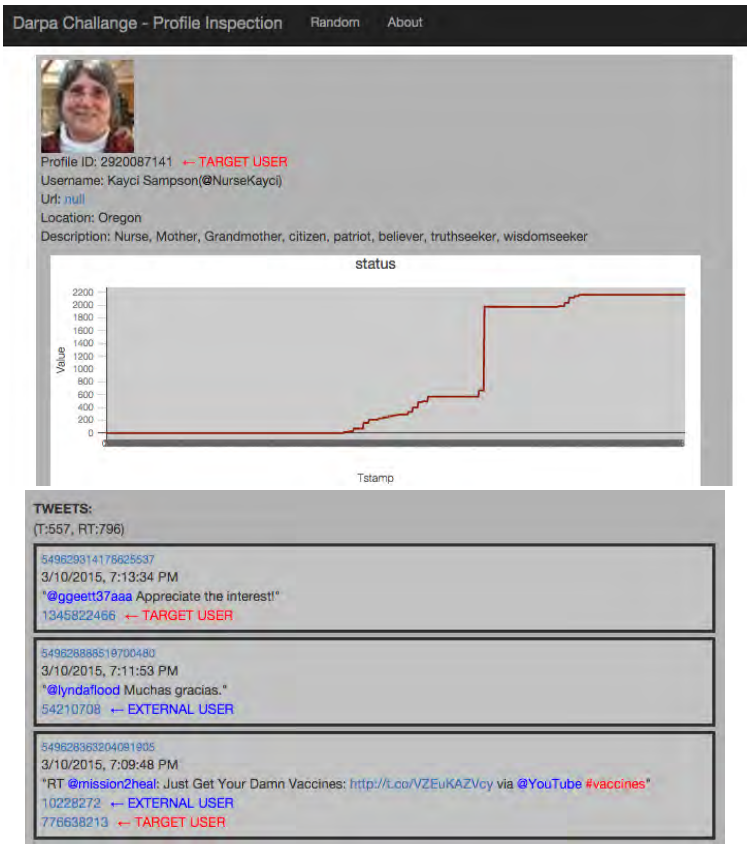



**Fig. 5.5.2.3**: Visualization and interactive data inspection interface.

| | **Feature description** |
|---|---|
| User metadata | # of posts/retweets/replies/mentions, min. # of status, # of friends/followers, GPS coordinate availability |
| Content | raw counts and fractions of pro- and anti-vaccine related content, hashtag/mention/url entropy |
| Sentiment | (a) happiness, (b) ANEW, (c) OpinionFinder, and (d) emoticon scores for each tweet and retweet |
| Network | retweet and mention network avg. clustering, core number, in-, out-degree centralities |
| Temporal | Signal-to-Noise ratio of user meta-data changes over time, max-min of these values and entropy of daily activity patterns |

**Table 5.5.2.1**: Classes of features to describe users profiles.

## 5.6 Computational fact-checking from knowledge bases

Coordinated efforts to spread information are especially a concern when less than transparent methods of promotion are adopted, when obscure or hidden interests lie behind the effort, and, of course, when the information being spread is unreliable. Online communication platforms, in particular social media, have created a situation in which the proverbial lie "can travel the world before the truth can get its boots on." Misinformation, astroturf, spam, and outright fraud have become widespread.

However, under certain conditions, reliable knowledge transmission can take place online. For example, Wikipedia, the crowd-sourced online encyclopedia, has been shown to be nearly as reliable as traditional encyclopedias, even though it covers many more topics. This motivated our attempt at leveraging any collection of factual human knowledge, such as Wikipedia, for automatic fact checking.

### *5.6.1 Introduction*

We focused on the simplest kind of factual statements that can be verified: let a *statement of fact* be represented by a subject-predicate-object triple, e.g., ("Socrates," "is a," "person"). A set of such triples can be combined to produce a *knowledge graph* (KG), where nodes denote *entities* (i.e., subjects or objects of statements), and edges denote predicates connecting the subject and object of a statement.

In a KG distinct paths between the same subject and object typically provide different factual support for the statement those nodes represent, even if the paths contain the same number of intermediate nodes. For example, paths that contain generic entities, such as "United States" or "Male," provide weaker support because these nodes link to many entities and thus yield little specific information. Conversely, paths comprised of very specific entities, such as "positronic flux capacitor" or "terminal deoxynucleotidyl transferase," provide stronger support. A fundamental insight that underpins our approach is that the definition of path length used for fact checking should account for such information-theoretic considerations.

To test our method [29] we use the DBpedia database (http://dbpedia.org) which consists of all factual statements extracted from Wikipedia's "infoboxes". From this data we build the large-scale *Wikipedia Knowledge Graph* (WKG), with 3 million entity nodes linked by approximately 23 million edges, see Fig. 5.6.1.1(a). These provide the most factual and uncontroversial information of Wikipedia.

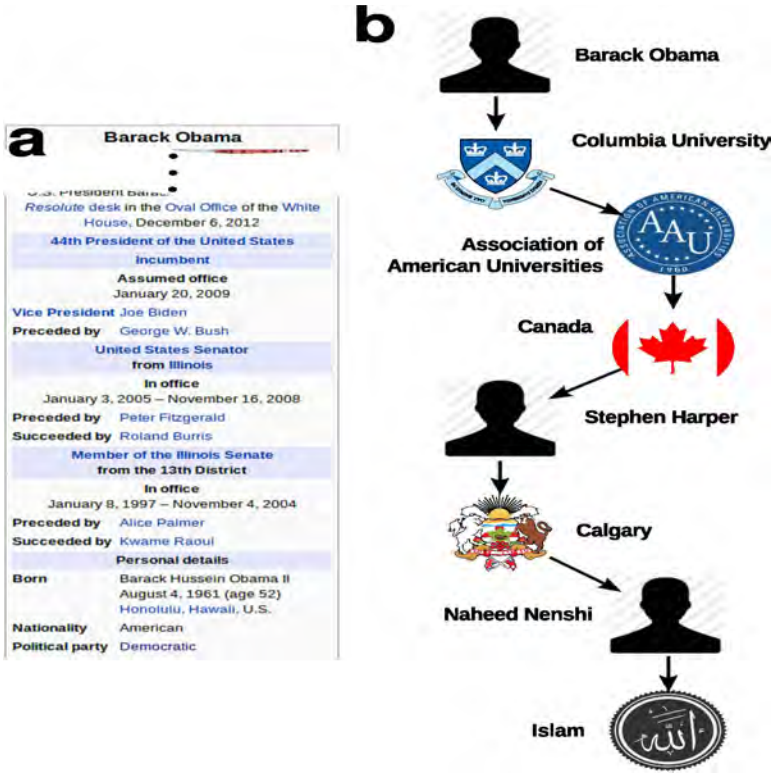



**Fig. 5.6.1.1**: Using Wikipedia to fact-check statements. (**a**) To populate the knowledge graph with facts we use structured information contained in the 'info-boxes' of Wikipedia articles (in the figure, the info-box of the article about Barack Obama). (**b**) In the diagram we plot the shortest path returned by our method for the statement "Barack Obama is a Muslim." The path traverses high-degree nodes representing generic entities, such as Canada, and is assigned a low truth-value.

### *5.6.2 Methods*

Let the WKG be an undirected graph $G = (V, E)$ where $V$ is a set of concept nodes and $E$ is a set of predicate edges. Two nodes $v, w \in V$ are said to be *adjacent* if there is an edge between them $(v, w) \in E$. They are said to be *connected* if there a sequence of $n \geq 2$ nodes $v = v_1, v_2, \ldots v_n = w$, such that, for $i = 1, \ldots, n - 1$ the nodes $v_i$ and $v_{i+1}$ are adjacent. The *transitive closure* of $G$ is $G^* = (V, E^*)$ where the set of edges is closed under adjacency, that is, two nodes are adjacent in $G^*$ *iff* they are connected in $G$ via at least one path. This standard notion of closure has been extended to weighted graphs, allowing adjacency to be generalized by measures of path length [Simas and Rocha (2014)], such as the semantic proximity for the WKG we introduce next.

The truth value $\tau(e) \in [0, 1]$ of a new statement $e = (s, p, o)$ is derived from a transitive closure of the WKG. More specifically, the truth value is obtained via a path evaluation function: $\tau(e) = \max \mathrm{W}(P_{s,o})$. This function maps the set of possible paths connecting $s$ and $o$ to a truth value $\tau$. A path has the form $P_{s,o} = v_1 v_2 \ldots v_n$, where $v_i$ is an entity node, $(v_i, v_{i+1})$

is an edge, $n$ is the path length measured by the number of its constituent nodes, $v_1 = s$, and $v_n = o$. Various characteristics of a path can be taken as evidence in support of the truth value of $e$. Here we use the *generality* of the entities along a path as a measure of its length, which is in turn aggregated to define a *semantic proximity*:

$$W(P_{s,o}) = W(v_1 \ldots v_n) = [1 + \sum_{i=2,n-1} \log k(v_i)]^{-1},$$

where $k(v)$ is the degree of entity $v$, i.e., the number of WKG statements in which it participates; it therefore measures the generality of an entity. If $e$ is already present in the WKG (i.e., there is an edge between $s$ and $o$), it should obviously be assigned maximum truth. In fact $W = 1$ when $n = 2$ because there are no intermediate nodes. Otherwise an indirect path of length $n > 2$ may be found via other nodes. The truth value $\tau(e)$ maximizes the semantic proximity defined by Eq. 1, which is equivalent to finding the shortest path between $s$ and $o$, or the one that provides the maximum information content in the WKG. The transitive closure of weighted graphs equivalent to finding the shortest paths between every pair of nodes is also known as the *metric closure*. This approach is also related to the Path Ranking Algorithm[6], except that here we use the shortest path (equivalent to maximum probability) rather than combining a sample of bounded-length paths in a learning framework.

### *5.6.3 Validation*

We test our fact-checking method on tasks of increasing difficulty, and begin by considering simple factual statements in four subject areas related to entertainment, history, and geography. We evaluate statements of the form "$d_i$ directed $m_j$," "$p_i$ was married to $s_j$," and "$c_i$ is the capital of $r_j$," where $d_i$ is a director, $m_j$ is a movie, $p_i$ is a US president, $s_j$ is the spouse of a US president, $c_i$ is a city, and $r_j$ is a country or US state. By considering all combinations of subjects and objects in these classes, we obtain matrices of statements. Many of them, such as "Rome is the capital of India," are false. Others, such as "Rome is the capital of Italy," are true. To prevent the task from being trivially easy, we remove any edges that represent true statements in our test set from the graph. Fig. 5.6.3.1 shows the matrices obtained by running the fact checker on the factual statements. Let $e$ and $e'$ be a true and false statement, respectively, from any of the four subject areas. To show that our fact checker is able to correctly discriminate between true and false statements with high accuracy, we estimate the probability that $\tau(e) > \tau(e')$. To do so we plot the ROC curve of the classifier since the area under the ROC curve is equivalent to this probability. With this method we estimate that, in the four subject areas, true statements are assigned higher truth-values than false ones with probability 95%, 98%, 61%, and 95%, respectively.

These findings represent a first step toward scalable computational fact-checking methods that may one day mitigate the spread of harmful misinformation. Further work is needed before present methods could be reliably applied in the wild.

---

[6] Lao, Ni, Tom Mitchell, and William W. Cohen. 2011. "Random Walk Inference and Learning in a Large Scale Knowledge Base." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 529–539. EMNLP '11. Stroudsburg, PA, USA: Association for Computational Linguistics.

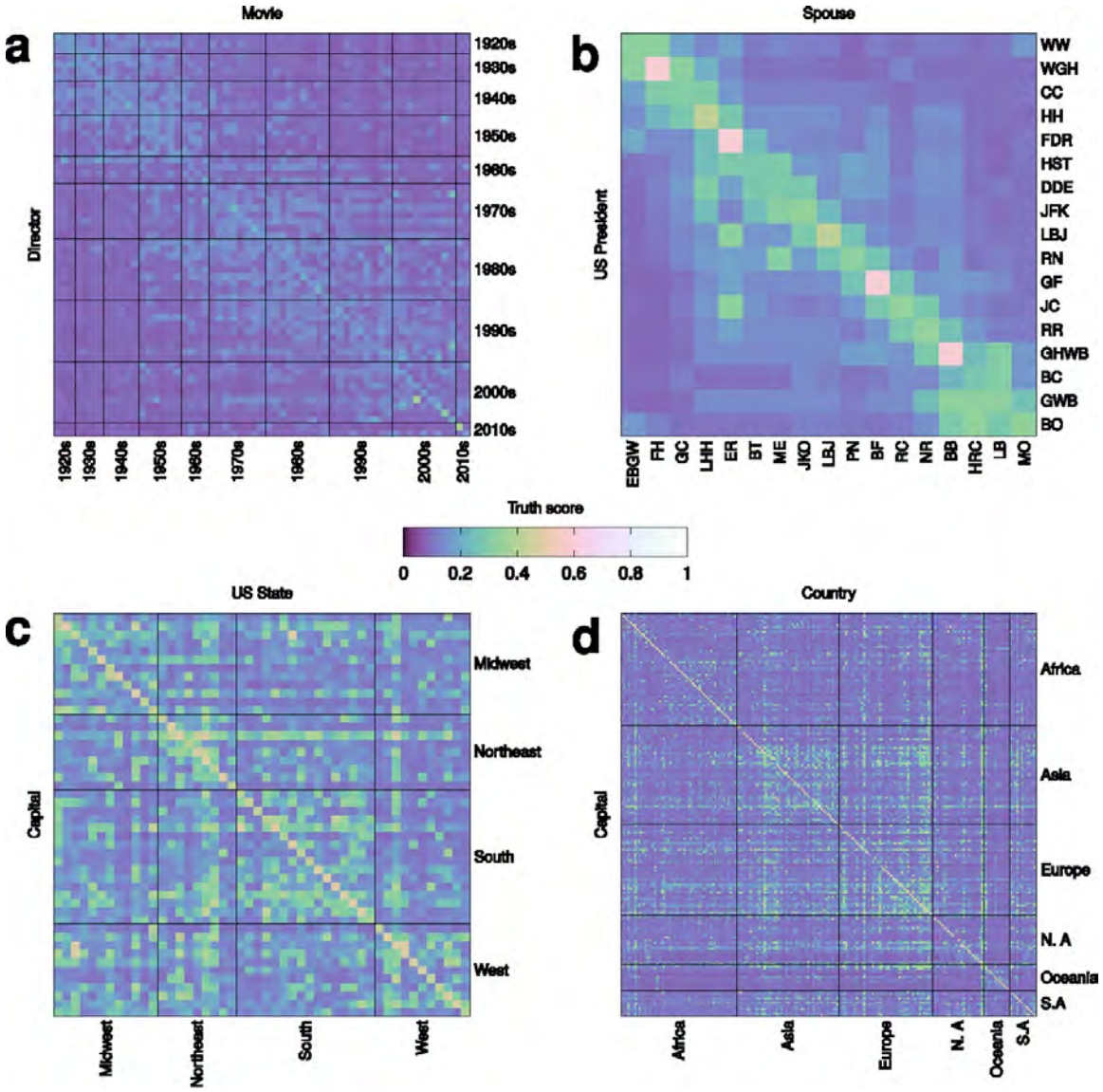



**Fig. 5.6.3.1**: Automatic truth assessments for simple factual statements. In each confusion matrix, rows represent subjects and columns represent objects. The diagonals represent true statements. Higher truth-values are mapped to colors of increasing intensity. (**a**) Films winning the Oscar for Best Movie and their directors, grouped by decade of award. (**b**) US presidents and their spouses, denoted by initials. (**c**) US states and their capitals, grouped by US Census Bureau-designated regions. (**d**) World countries and their capitals, grouped by continent.

## 5.7 Detection and classification of persuasion campaigns on Twitter

The major focus of ATL project was to study temporal behavior of information cascades by tracking the feature vectors representing information diffusion. We generate multi-dimensional time series or cascade trajectory describing individual cascade evolution.

We think that cascade trajectories can represent different classes of conversation patterns

that occur in online social media. We assume that a cascade signature matching process similar to anomaly detection can detect orchestrated deception campaigns.

### 5.7.1 Multi-dimensional time series analysis with SAX-VSM technology

In the design of our framework we had to take into account three important factors: (i) Multi-dimensional time series raw data can grow in size easily over gigabytes; this aspect explicitly affects the scalability of the algorithms. (ii) In addition to the problem of large volume of data, most classic machine learning algorithms do not work well on time series raw data due to their unique structure; in our scenario, time series have a very high dimensionality, high feature correlation, and a large amount of noise, which present a difficult challenge in time series data mining tasks. (iii) As a result, many time series algorithms instead of operating on the original "raw" data, better operate on higher-level representation of such data.

We proposed a novel method for temporal data analysis and classification, called SAX-VSM, which is based on two existing techniques namely, SAX (Symbolic Aggregate approXimation) and VSM (Vector Space Model). The SAX-VSM algorithm demonstrates a high accuracy performance, learns efficiently from a small training set, and has a low computational complexity.

#### *SAX: Symbolic Aggregate ApproXimation*

The basic idea of Symbolic Aggregate Approximation, SAX is to convert the data into a discrete format, with a small alphabet size. To convert a time series into symbols, first the time series is normalized, and then two steps of discretization are performed. In details, a time series is initially transformed using Piecewise Aggregate Approximation (PAA). This method simply approximates a time series by dividing it into equal-length segments and recording the mean value of the data points that fall within each segment.
Next, to convert the PAA values to symbols, user determines the breakpoints that divide the distribution space into $\alpha$ equiprobable regions, where $\alpha$ is the alphabet size specified by the user. The PAA coefficients can then be easily mapped to the symbols corresponding to the regions in which they reside.
Figure 5.7.1.1 shows an example of a time series being converted to string "baabccbc". Note that the general shape of the time series is still preserved, in spite of the massive amount of dimensionality reduction.

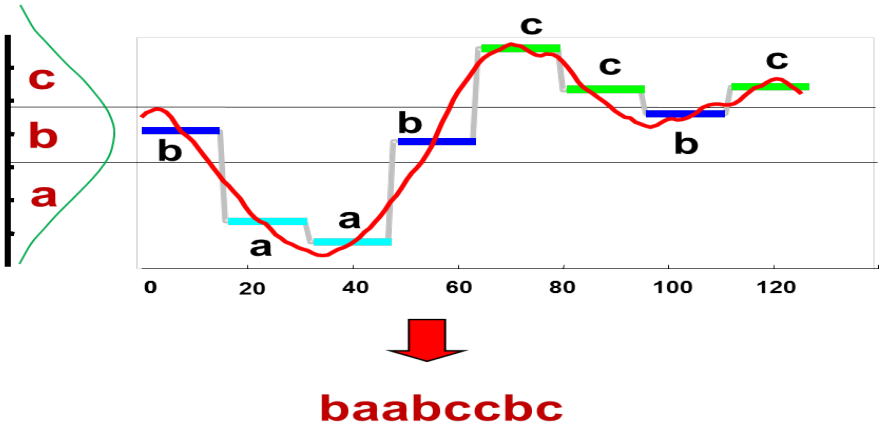



**Fig. 5.7.1.1:** A visualization of the SAX dimensionality reduction technique. A time series (red line) is discretized first by a PAA procedure (N = 8) and then, using breakpoints of arbitrary length, it is mapped into the word "baabccbc" using an alphabet size of 3.

### *Vector Space Model, VSM*

The second component of SAX-VSM technique is known in Information Retrieval a Vector Space Model, VSM. In order to build SAX words vocabularies of long time series we use a sliding window technique to convert a time series into the set of SAX words. By sliding a window across time series, extracting subsequences, converting them to SAX words, and placing these words into an unordered collection, we obtain the "Bag of Words" representation of the original time series. Each row of the constructed matrix (Bag of Words) represents a SAX word and corresponding frequency of that word generated by sliding window procedure (see Fig. 5.7.1.2). Following the common Information Retrieval workflow, we employ the TF*IDF weighting scheme for each element of this matrix in order to transform a frequency value into the weight coefficient.

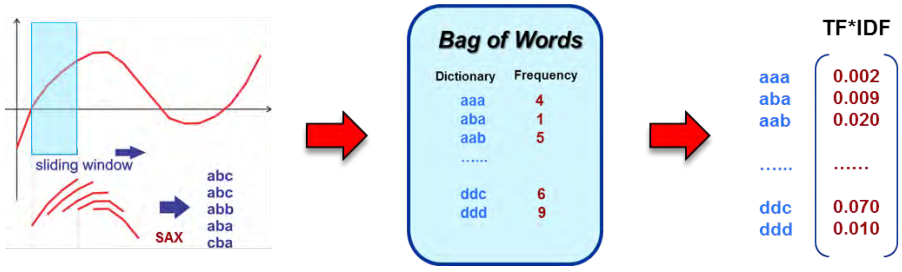



**Figure 5.7.1.2**: By sliding a window across time series, extracting subsequences, converting them to SAX words, and placing these words into an unordered collection, we obtain the bag of words representation of the original time series. Next, TF*IDF statistics is computed resulting in a single weight vector.

### SAX-VSM classification procedure

Similar to other classification techniques, SAX-VSM consists of two parts - the training phase and the classification procedure. An overview of the SAX-VSM algorithm is shown in Figure 5.6.3. In the training phase, all labeled time series from N training classes are transformed into symbolic representation, and the algorithm generates N TF*IDF weight vectors representing N training classes (see Fig. 5.7.1.3).

In the classification phase, an unlabeled time series is converted into a term frequency vector and assigned to the class whose TF*IDF weight vector has a maximal cosine similarity.

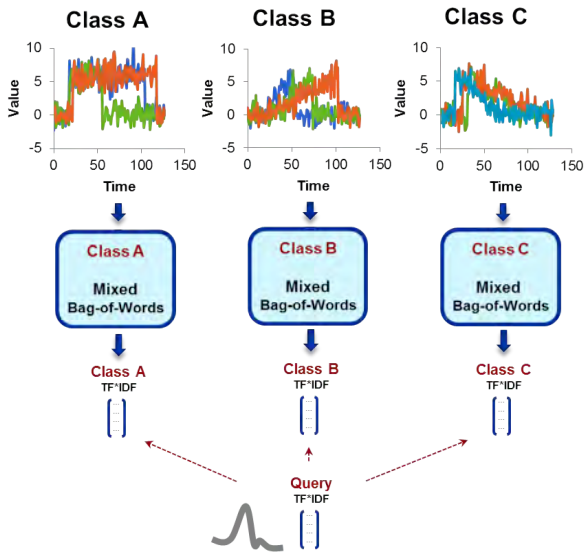
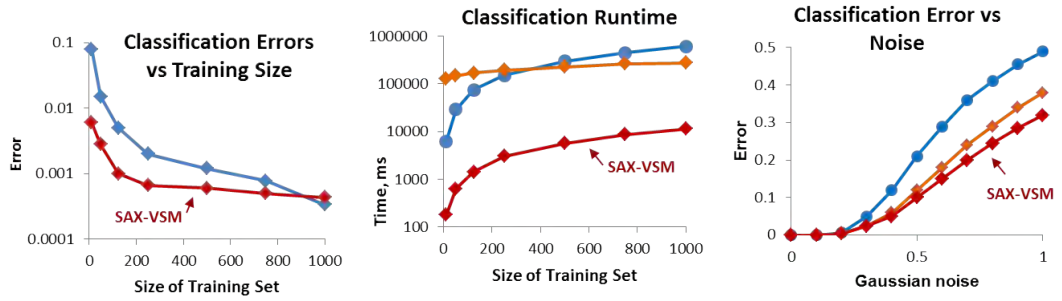



**Figure 5.7.1.3**: An overview of SAX-VSM algorithm: at first, all labeled time series from each class are converted into a single bag of words using SAX; secondly, TF*IDF statistics is computed resulting in a single weight vector per training class. For classification, an unlabeled time series is converted into a term frequency vector and assigned a label of a weight vector, which yields a maximal cosine similarity value.

### *SAX-VSM characteristics: accuracy, performance and tolerance to noisy data*

We used well-known synthetic CBF test in order to investigate and compare the performance of SAX-VSM and 1NN Euclidean classifier on increasingly large datasets (See Fig. 5.7.1.4). Detailed analysis of SAX-VSM performance and comparison with other temporal data classification techniques is described in detail in our original paper.




(a) (b) (c)

**Figure 5.7.1.4**: SAX-VSM algorithm characteristics: (a) - Comparison of classification precision and run time of SAX-VSM (red) and 1NN Euclidean classifier (blue) on CBF data. SAX-VSM performs significantly better with limited amount of training samples. (b) - While SAX-VSM is faster in time series classification, its performance is comparable to 1NN Euclidean classifier when training time is accounted for. (c) -SAX-VSM increasingly outperforms 1NN Euclidean classifier with the growth of a noise level.

The unique characteristics of SAX-VSM, such as high classification accuracy, learning efficiency and a low computational complexity suggested using SAX-VSM for the goal of current research.

### *Extending SAX-VSM for n-dimensional case*

The described SAX-VSM algorithm can be extended easily to n-dimensional case. Each dimension of multi-dimensional time series (trajectory) can be processed independently in terms of calculating corresponding Bags-of-Words and TF*IDF weight vectors for each dimension. To compare two trajectories, ***A*** and ***B***, cosine similarities along each dimension can be calculated in the same way as it was done in one-dimensional case and then total similarity of trajectories can be estimated by combining similarities along all directions:

$$sim(A,B) = \sqrt{\frac{\sum_{i=1}^{n} sim(A,B)_i^2}{n}}$$

## 5.7.2 SAX-VSM for Twitter data classification

In collaboration with the IU team, we focused on two classes of information diffusion on Twitter: advertisement campaigns defined as promoted content on Twitter as opposed to non-promoted trending topics. This choice allows generating large-scale data that do not require human efforts for labeling.

### *Experiments and classification results*

The dataset consisted of 76 promoted and 853 non-promoted trends. 224 features selected among network features, event-interval features and user characteristics characterized each time series.

In our classification tests we used a Leave-One-Out Cross-Validation (LOOCV) approach: we systematically applied our multi-dimensional version of SAX-VSM classifier for each example using the rest of the sample for training.
Initially the feature selection procedure was organized in the following way: we pipelined a Monte Carlo random search of feature combinations and LOOCV test. To reduce the search in potentially very large combinatorial space, we ranked individually all 224 features by their classification ability and then limited the search space by using only the 60 top features. We achieved good results in classification quality, keeping only 12 features and randomly testing possible combinations of 12 from the top 60 available features. The best features found this way are arranged according to their descending ranks and shown in Table 5.7.2.1. Together they produce classification accuracy of 97%.

| Feature Name | Description |
|---|---|
| hashtagN_degree_skewness | Skewness of degree distribution (hashtag network) |
| hashtagN_CC_min | Min. clustering coeff. (hashtag network) |
| Frequency | Volume of tweets |
| mentionN_LCC_mean_shortest_path | Mean shortest-path (LCC) of the mention network |
| retweetN_density | Density of the retweet network |
| event_interval_mean | Mean of distribution of tweets time intervals |
| hashtagN_degree_entropy | Entropy of degree distribution (hashtag network) |
| event_retweet_interval_kurtosis | Kurtosis of distribution of retweets time intervals |
| user_favourites_count_min | Min. of distribution of favorite tweets |
| event_mention_interval_entropy | Entropy of distribution of mentions time intervals |
| event_mention_interval_std | Std. dev. of distribution of mentions time intervals |
| event_interval_skewness | Skewness of distribution of tweets time intervals |

**Table 5.7.2.1:** Set of features giving best classification results

To evaluate the performance of binary classifying systems like our SAX-VSM procedure, it is a common practice to calculate a Receiver Operating Characteristic (ROC), or ROC curve. By plotting the true positive rate vs. the false positive rate at various threshold settings and measuring the area under the ROC curve (AUC), we get another evaluation of classifier accuracy. In Figure 5.7.2.1, the plot of the ROC is shown for the case of 12 features included in Table 5.7.2.1.

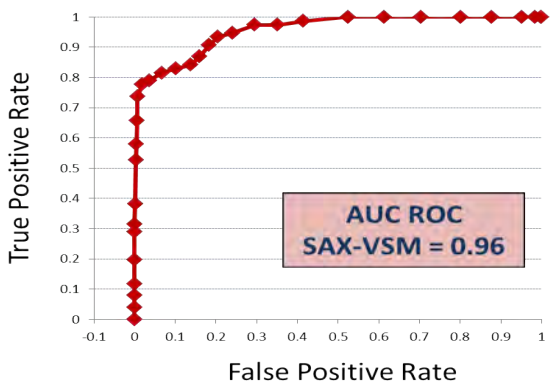



**Figure 5.7.2.1:** ROC curve for the classification experiment

### 5.7.3 Improving feature selection

We implemented and experimented with a classical Forward Selection (FS) algorithm and slightly modified Restricted Forward Selection (RFS). We chose True Positive Rate as the utility function. Our experiments demonstrated that both algorithms, FS and RFS, yield very good results. Below is the result obtained by using the FS algorithm and optimization performed at the trending phase. The optimization included all network and event features with our usual parameter set: PAA = 4, alphabet =5, SAX window = 70.

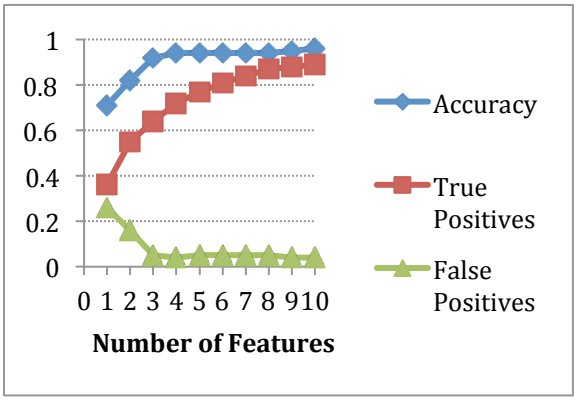



**Figure 5.7.3.1**. Improving detection accuracy during the feature selection process.

| Top 10 features |
|---|
| **mentionN_degree_mean** |
| **retweetN_degree_std** |
| **hashtagN_CC_max** |

| |
|---|
| **retweetN_CC_kurtosis** |
| **mentionN_nodes** |
| **hashtagN_edges** |
| **hashtagN_CC_mean** |
| **event_retweet_interval_entropy** |
| **mentionN_out-degree_std** |
| **event_interval_mean** |

**Table 5.7.3.1**: The 10 best features obtained by running a feature selection process, FS, described above. The features are arranged according to their descending ranks.

### 5.7.4 Conclusions

ATL demonstrated that without any content analysis of topics on Twitter™, by monitoring only temporal traces of topological characteristics of users' networks with twitting temporal activity, it is possible to distinguish two types of topics on Twitter™, promoted or advertisement campaigns and non-promoted or naturally trending topics. We presented experimental results of applying our SAX-VSM classification technique of multidimensional time series to achieve high detection accuracy on Twitter™ data. Our results suggest that social streams can be monitored effectively almost in a real time and some abnormal activity can be detected by analyzing temporal evolution of social networks.

## 5.8 Predicting bursts and rumors in social media

### 5.8.1 Introduction

Co-PI Dr. Qiaozhu Mei, his graduate student assistants, and a postdoctoral researcher have been conducting research supported by this project. Our goal in this project is to detect social media rumors early and to understand their diffusion patterns.

Rumors, which are intentionally used in many persuasion campaigns, can be extremely harmful. Especially in social media, a single tweet containing a false rumor has the potential of deceiving millions of people and hurting businesses in minutes. For instance, a Tweet posted by an hacked Twitter account of the Associated Press (AP) on April 23rd, 2013, which was about an alleged bombing in the White House, made the stock market take an instant nosedive. A steep recovery followed when the hacking became apparent. These changes between the spread and the correction of the rumor took just 15 minutes. Similar incidents could potentially have a huge tangible impact.

Rumors usually contain a piece of controversial and factual statement on topics of general interest. A social context for rumors to spread widely is when people have unmet information needs related to the topic. We worked at identifying conversations expressing an information need [25]. Specifically, we built a classifier that identified Tweets containing specific information needs, i.e., Tweets that are real

questions, with 85% accuracy. We conducted a longitudinal comparative analysis of these questions and the overall Twitter corpus. We showed that questions users ask on Twitter can be used as a signal to detect rumors (controversial factual statements). We observed that the uncertainty about the veracity of a rumor triggers questions from common Twitter users. On being exposed to an unverified rumor, many people immediately seek additional information to verify the truth-value of the rumor. Based on this observation, we developed an efficient algorithm to detect emerging rumors as early as possible, and certainly before they enjoy a peak in popularity. At the base of our method is the developed framework to filter Tweets containing enquiries [15]. The results showed that our framework could detect rumors from raw Tweet stream at a high precision, and several hours before they have been clarified or corrected by either authoritative sources or the crowd.

The online rumor detector is the first step to identify potential rumors and prevent the harmful ones from spreading. Once a candidate rumor is detected, the next task is to gather more information about the rumor, possibly by retrieving every Tweet spreading or correcting it. In this way, the domain experts can investigate the candidate rumor, evaluate its spreading behavior, and decide on its truth-value. We developed a retrieval system that can retrieve all other Tweets relevant to the detected rumors with limited use of human judgments [12]. We also built a visualization tool for domain experts to analyze how a rumor spreads based on the analysis of the audience of the rumor and its corrections [16].

In the rest of this section, we will summarize particular components of our work. We begin with analyzing the user's information-seeking behavior, followed by the discussion of the early rumor detection algorithm. We then present a user-in-the-loop, high precision and high recall retrieval system to detect individual Tweets discussing a specific rumor. We will also present a tool that helps people visualize and evaluate the spread of rumors. Finally, we briefly discuss our approach to detecting influence bots in the SMISC challenge, which we participated in with our collaborators at Indiana University.

### 5.8.2 Detection and Analysis of Questions on Twitter

Being exposed to a rumor, many people immediately seek additional information to verify the truth-value of the rumor. In other words, the uncertainty about the veracity of a rumor drives people to ask questions. By studying users' information seeking behavior, mainly questions, we may find important signals to identify controversial persuasion campaigns in their early stages.

Conventional studies of online information seeking behavior usually focused on the use of search engines or question answering (Q&A) websites. In this study, we proposed to extract and analyze questions from billions of online conversations collected via Twitter, which was published at the international World Wide Web conference in 2013 [25]. We trained a text classifier to detect information-seeking questions on Twitter, which achieved an accuracy of 86.6%. We did a comprehensive analysis of the types of questions we extracted. We found that the questions being asked on Twitter were substantially different from the topics tweeted in general. Information needs detected on Twitter had a considerable power of predicting the trends of Google queries. We also conducted a longitudinal analysis of the volume,

spikes, and entropy of questions on Twitter to study the impact of real world events and user behavioral patterns on social platforms.

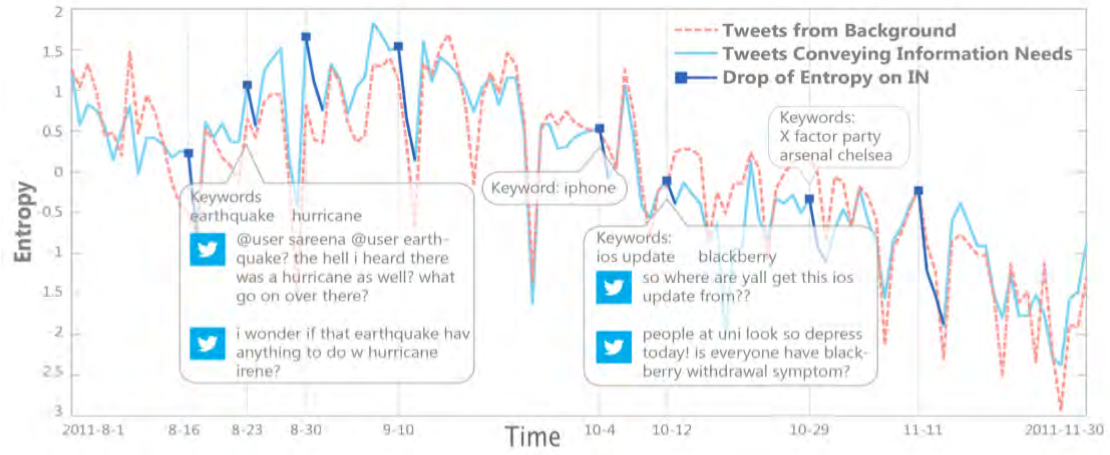



**Figure 5.8.2.1**: Longitudinal analysis showing examples of tweets with information need [25].

Figure 5.8.2.1 above shows an example of an entropy analysis on the questions being asked on Twitter. It plots the entropy of the language models of all information needs, and of all tweets in the general discussion over time. We mark several points in the time series each of which presents a sudden entropy drop the next day, which indicates a concentration of the topics being discussed/asked. We extracted the keywords that are significantly overrepresented on the day after each marked point, which gave us a basic idea about the topics that triggered the concentration. These topics typically corresponds to surprising events occurred at that time, including rumors.

### 5.8.3 Early Detection of Rumors in Social Media

Our analysis of the information seeking behavior, as described above, showed that rumors triggers questions asked by Twitter users and therefore can be potentially detected by tracking these questions. Indeed, we observed that many users who were exposed to a rumor would seek more information about a rumor before deciding whether to believe it, spread it, or debunk it. Some of these enquiries happened over Twitter. Based on this observation, we designed a rumor detection framework that can efficiently detect rumors at a very early stage of their lifecycle. Potential rumor statements were identified based on the kind of enquiries they generated. For example, Table 5.8.3.1 shows some enquiry tweets that were sent out within 60 seconds of the tweet from the hacked Twitter account of the Associated Press in April 2013 about two explosions in the White House.

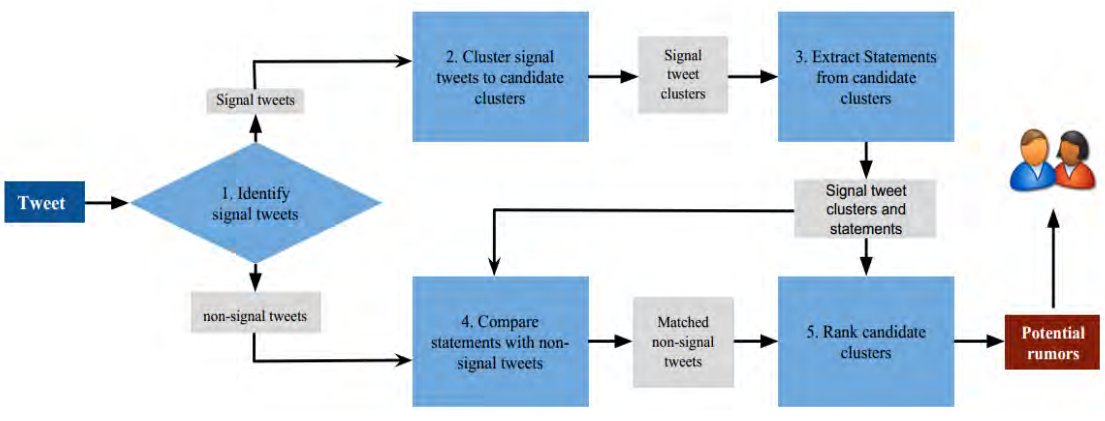
| |
|---|
| Oh my god is this real? RT @AP: Breaking: Two Explosions in the White House and Barack Obama is injured. |
| Is this true? Or hacked account? RT @AP Breaking: Two Explosions in |

| |
|---|
| the White House and Barack Obama is injured. |
| Is this real or hacked? RT @AP: Breaking: Two Explosions in the White House and Barack Obama is injured. |
| How does this happen? #hackers RT @user: RT @AP: Breaking: Two Explosions in the White House and Barack Obama is injured. |
| Is this legit? RT @AP Breaking: Two Explosions in the White House and Barack Obama is injured. |

**Table 5.8.3.1**: Examples of enquiry tweets about the rumor of explosions in the White House

The framework of our rumor detection algorithm is shown in Figure 5.8.3.1. We first identify signal tweets using a set of regular expressions, e.g., "is this true?", "what?", "this is not true", etc. in order to select only those Tweets that contain skeptical enquiries: i.e. verification questions and corrections. Then, we cluster the signal tweets based on overlapping content in the Tweets and analyze the content of each cluster to determine a single summary statement for the cluster. We then capture all non-signal Tweets that match any of the cluster summary statements to identify candidate rumor clusters. Finally, we rank the candidate rumor clusters by the likelihood that their statements are rumors using statistical features independent of the statements' content.

To evaluate our approach to detect rumors, we conducted experiments over two Twitter collections. One was focused on the Boston marathon bombing, a major newsworthy event in April 2013. The other was the "background" collection consisting of a random sample of Tweets from a month with no significant newsworthy event. We first compared our method against two baseline methods: (a) detecting bursting events, and (b) tracking trending social memes. The results over the Boston bombing collection showed that more than half of the statements outputs by our approach were rumors, while the two baseline methods obtained a much lower accuracy. We also compared our approach with another method that tracks correction patterns, such as presence of keywords like "rumor", "debunk", etc. Our approach returned more rumors with a comparative accuracy. We also showed that our approach could detect rumors about three hours earlier than any of



the baseline methods.

**Figure 5.8.3.1**: The procedure of real-time rumor detection (Figure 1 in Zhao et al. [15]).

The experiments on the background collection showed that our approach effectively and efficiently detected rumors. With a 72-core Hadoop cluster, in about half an hour we were able to monitor and process 10% of all the tweets posted in one day on Twitter. Out of 50-candidate statements output by the approach, about one third were real rumors, and about 70% of the ten top ranked candidates were real rumors. For further details, please refer to our paper [15], which was published at the international World Wide Web conference in 2015. Figure 5.8.3.2 shows the lifespan of rumors we detected in the Boston bombing collection.

### 5.8.4 Rumor Retrieval and Visualization

Next, we aim to extract meaningful signals for detecting rumors and other types of persuasion campaigns, and to analyze and model the diffusions of the rumors and the corresponding corrections. To do these, we needed to identify **all** Tweets relevant to the rumors we detected, including those using different words and expressions but conveying the same rumor.

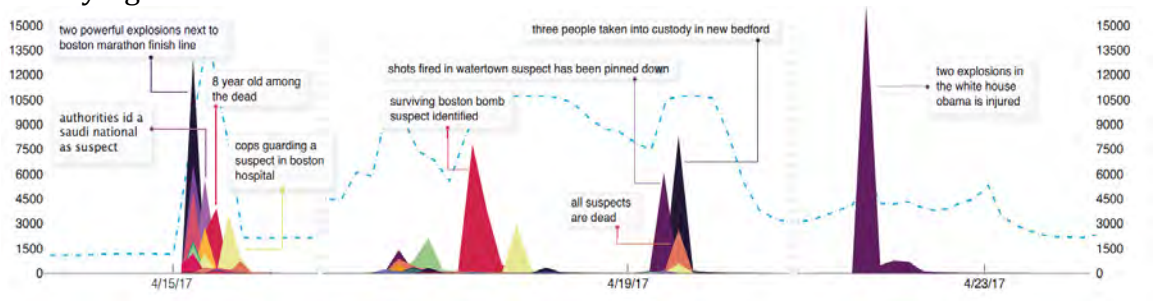



**Figure 5.8.2.2:** Detect and track rumors rom Boston marathon explosion (Figure 5 in Zhao et al. [15]).

To address this, we proposed a novel framework of retrieval techniques that is particularly useful for maximizing the recall of relevant results without compromising the precision. It could be used to effectively retrieve relevant Tweets given a rumor statement. This new framework features a ReQ-ReC (ReQuery-ReClassify) process, a double-loop retrieval system that combines iterative expansion of a query set with iterative refinements of a classifier. The flowchart of this framework is shown in Figure 5.8.4.1.

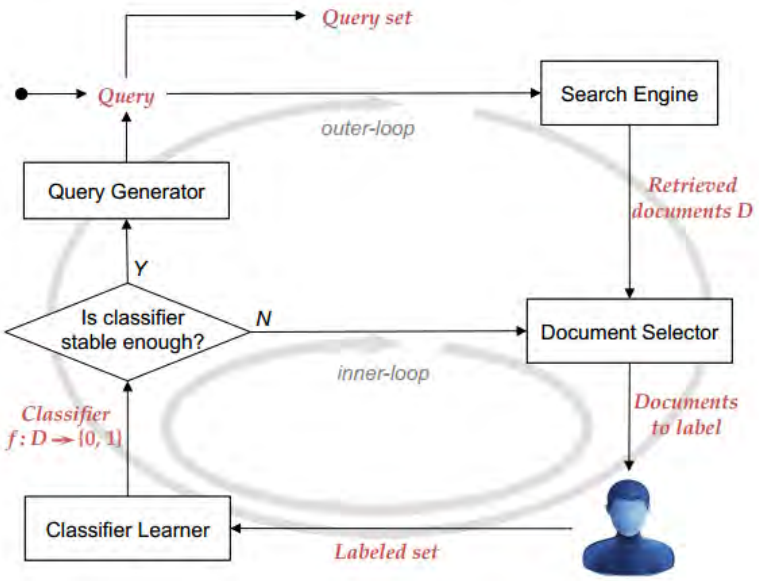



**Figure 5.8.3.1:** ReQ-ReC framework (Figure 1 in Li et al. [12]).

This new framework permits a separation of labor: the query generator's task is to enhance the recall while the classifier's one is to maximize the precision on the results retrieved by any of the queries. The overall process alternates between the query expansion cycle (the outer loop to increase recall) and the classifier refinement cycle (the inner loop to increase precision). The separation of the two roles allows the query enhancement process to be more aggressive in exploring new query suggestions. Our results are published at the ACM SIGIR Conference on Research and development in information retrieval in 2014.

We evaluated the framework on four large datasets provided by the TREC. Our experiments show that the separation of roles significantly outperforms other relevance feedback methods that rely on a single ranking function to balance precision and recall. On average, the new framework yields a 20% to 30% improvement of recall-oriented retrieval performance (measured by R-precision or mean average precision) on very large microblog data sets. Additional details can be found in (Li et al., [12]). The initial versions of the framework were used to participate in the microblog track of the TREC 2013 and achieved the top rankings among more than 70 submissions from 20 participating teams [16].

We also developed a tool to analyze and visualize the diffusion of rumors [27]. The basic idea is to partition the audience of a rumor and its corrections into several states and model their transition between states. The RumorLens tool consists of two stages: a pre-computation stage and a visualization stage. The first stage takes a dataset of Tweets that have been tagged as propagating or correcting a rumor, as well as the social network of the propagating users, and calculates the marginal impact of each tweet. The second stage helps visualize the result of the first stage. The tool could be used to better understand the diffusion of rumors and find other interesting and meaningful facts about them. Our results have been presented at the international conference of weblog and social media in 2015.

### 5.8.5 Influence detection prediction strategy: Multi-arm bandit

We also participated in the SMISC influence detection challenge as a team in collaboration with the Indiana University. Two specific aspects of the challenge motivated us to adopt an online prediction strategy to detect bots. First, the data was progressively released to the participating teams, and not made available all at once. Second, during the live challenge, once a guess was submitted, the participating teams received immediate feedback on the correctness of that guess. Such a setting closely fits an online prediction scenario, where the learner observes a data stream and tries to make increasingly accurate guesses based on receiving feedback on its guesses and updating its internal belief.

We adopted the multi-arm bandit paradigm as our online prediction strategy. We initialized the "arms" with a set of binary independent classifiers, and used a variant of the hedge algorithm[7] as the meta-learning strategy which decided which arm to pull next. Specifically, each binary classifier based its prediction on different aspects of the user profile in such a way that when combined, the classifiers could complement each other in detecting different variants of bots.

Each user account got a prediction score between [0,1] assigned by each arm. A higher score indicates that the classifier assigns higher likelihood for the user account to be a bot. The meta-learner, which executed the variant of the hedge algorithm, initially assigned uniform weights to each arm, and then used a multiplicative scheme to update the arm weights. After each round, the meta-learner produced a final "bot score" for each user account as the weighted average of prediction scores output by each arm. Finally, it selected the account with the highest bot score as the next guess. On receiving the feedback score $x$ (which could be positive or negative), the weight of all classifiers were multiplied by a factor of $e^{(x.f_j)}$, where $f_j$ is classifier $j$'s prediction score for the guessed account. This way, a classifier ("arm") would gradually gain weight if it accurately detected bots with high confidence, and gradually lose weight if it mislabeled normal accounts as bots with high confidence. Learning from our predictions in the SMISC challenge, we intend to continue working to tune our strategy to detect other persuasion campaigns.

### 5.8.6 Summary

Persuasion campaigns, especially the ones on controversial topics, usually use false rumors to achieve their objectives. Our work in this project aims to detect, retrieve and understand social media rumors. With our achievements, domain experts ahead of those rumors' outbursts and unpredictable consequences can take actions and decisions.

After analyzing user's information seeking behaviors, we discovered a set of signals that appear mostly at the early stage of rumors. We then take raw tweet stream as input, monitor signal Tweets, group them, and output potential rumor clusters. Next, by applying our ReQ-ReC retrieval system, we managed to retrieve relevant Tweets

[7] Yoav Freund and Robert E. Schapire, "A decision-theoretic generalization of online learning and an application to boosting," Journal of Computer and System Sciences 1997;55(1):119-139.

about the detected rumor with a high precision and a high recall. At last, besides the rumor statement and its tweets, domain experts will also see a highly descriptive visualization of its diffusion.

## 6 List of DESPIC publications


1. O. Varol and F. Menczer; *Connecting Dream Networks Across Cultures.* WWW Companion '14: Proceedings of the companion publication of the 23rd international conference on World Wide Web companion, 2014. http://dx.doi.org/10.1145/2567948.2579697

2. X. Gao, E. Roth, K. McKelvey, C. Davis, A. Younge, E. Ferrara, F. Menczer, and J. Qiu; *Supporting a Social Media Observatory with Customizable Index Structures-Architecture and Performance*. Book chapter in Cloud Computing for Data Intensive Applications, Springer, 2014 http://salsaproj.indiana.edu/IndexedHBase/paper_bookChapter.pdf

3. A. Nematzadeh, E. Ferrara, A. Flammini, and Y.Y. Ahn; *Optimal network clustering for information diffusion*. Physical Review Letters, 113, 088701, 2014. [Editor's pick] http://journals.aps.org/prl/abstract/10.1103/PhysRevLett.113.088701

4. M. JafariAsbagh, E. Ferrara, O. Varol, F. Menczer, and A. Flammini; *Clustering memes in social media streams*. Social Network Analysis and Mining Social Network Analysis and Mining 4 (237), 1-13, 2014. http://link.springer.com/article/10.1007/s13278-014-0237-x#page-1

5. E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini; *The rise of Social Bots.* Communications of the ACM – (to appear) http://arxiv.org/abs/1407.5225

6. O. Varol, E. Ferrara, C. Ogan, F. Menczer, and A. Flammini; *Evolution of online user behavior during a social upheaval*. ACM Web Science '14, pp. 81-90. ACM 2014 [Best Paper Award] http://dl.acm.org/citation.cfm?id=2615699

7. X. Gao and J. Qiu; *Supporting Queries and Analyses of Large-Scale Social Media Data with Customizable and Scalable Indexing Techniques over NoSQL Databases.* Proceedings of the 14th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid 2014). Chicago, IL, USA, May 26-29, 2014 http://mypage.iu.edu/~gao4/paper_ccgrid2014.pdf

8. X. Gao, J. Qiu; *Social Media Data Analysis with IndexedHBase and Iterative MapReduce*. Proceedings of the 6th Workshop on Many-Task Computing on Clouds, Grids, and Supercomputers (MTAGS 2013) at Super Computing 2013. Denver, CO, USA, November 17th, 2013. http://datasys.cs.iit.edu/events/MTAGS13/p07.pdf

9. L. Weng and F. Menczer; *Topicality and social impact: diverse messages but focused messangers*. PLoS ONE, 10(2): e0118410, 2015 http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0118410

10. P. Senin and S. Malinchik; *SAX-VSM: Interpretable Time Series Classification Using SAX and Vector Space Model*, Proc. of ICDM 2013, Dallas, Texas / December 7-10, 2013
http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=6729617

11. S. Malinchik; *Detection of Persuasion Campaigns on Twitter™ by SAX-VSM Technology*, Proc. of ICDS 2014, The Eighth International Conference on Digital Society, Barcelona, Spain, March 2014.
http://www.thinkmind.org/index.php?view=article&articleid=icds_2014_5_20_10080

12. C. Li, Y. Wang, P. Resnick, and Q. Mei; *ReQ-ReC: High Recall Retrieval with Query Pooling and Interactive Classification*. SIGIR '14 The 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, Gold Coast, QLD, Australia — July 06 - 11, 2014 http://www-personal.umich.edu/~qmei/pub/sigir2014-li.pdf

13. S. Kong, Q. Mei, L. Feng, F. Ye, and Z. Zhao; *Predicting Bursts and Popularity of Hashtags in Real-Time*. SIGIR '14 The 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, Gold Coast, QLD, Australia — July 06 - 11, 2014 http://www-personal.umich.edu/~qmei/pub/sigir2014-kong.pdf

14. X. Rong and Q. Mei; *Diffusion of innovations revisited: from social network to innovation network*. Proceedings of the 22nd ACM international conference on Conference on information & knowledge management, San Francisco, CA, USA — October 27 - November 01, 2013 http://www-personal.umich.edu/~qmei/pub/kdd2013-tang.pdf

15. Z. Zhao, P. Resnick and Q. Mei; *Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts*. WWW '15 Proceedings of the 24th International Conference on World Wide Web, 2015 http://dl.acm.org/citation.cfm?id=2741637

16. C. Li, Y. Wang, and Q. Mei; *A User-in-the-Loop Process for Investigational Search*: Foreseer in TREC 2013 Microblog Track, in Proceedings of the Twenty-Second Text REtrieval Conference (TREC 2013). http://trec.nist.gov/pubs/trec22/papers/foreseer-microblog.pdf

17. E. Ferrara, M. JafariAsbagh, O. Varol, V. Qazvinian, F. Menczer, and A. Flammini; *Clustering Memes in Social Media*. In Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM'13), 2013. IEEE/ACM
http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6785757

18. K. McKelvey and F. Menczer; *Design and prototyping of a social media observatory*. In Proceedings of the 22nd international conference on World Wide Web companion, pp. 1351-1358, May 13–17, 2013, Rio de Janeiro, Brazil. ACM 2013. http://dl.acm.org/citation.cfm?id=2488174

19. K. McKelvey and F. Menczer; *Interoperability of Social Media Observatories*. In Proceedings of the First International Workshop on Building Web Observatories. May 8, 2013, Paris, France. ACM http://cnets.indiana.edu/wp-content/uploads/websci13.pdf

20. K. R. McKelvey and F. Menczer; *Truthy: enabling the study of online social networks.* In Proceedings of the 2013 conference on Computer supported cooperative work companion, pp. 23-26, 2013. ACM 2013. http://dl.acm.org/citation.cfm?id=2441962

21. L. Weng, J. Ratkiewicz, N. Perra, B. Gonçalves, C. Castillo, F. Bonchi, R. Schifanella, F. Menczer, and A. Flammini; *The Role of Information Diffusion in the Evolution of Social Networks.* In *Proceedings of the 19th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, August 11-14, 2013, Chicago, USA. http://dl.acm.org/citation.cfm?id=2487607

22. M. D. Conover, C. Davis, E. Ferrara, K. McKelvey, F. Menczer, and A. Flammini*; The geospatial characteristics of a social movement communication network*. *PloS ONE*, 8(3), e55957, 2013
http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0055957

23. M. D. Conover, E. Ferrara, F. Menczer, and A. Flammini; *The Digital Evolution of Occupy Wall Street*. *PloS ONE*, 8(5), e64679, 2013. http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0064679

24. E. Ferrara, O. Varol, F. Menczer, and A. Flammini; *Traveling Trends: Social Butterflies or Frequent Fliers?* In Proceedings of the ACM Conference on Online Social Networks (COSN), 2013. ACM 978-1-4503-2084-9/13/10 http://dl.acm.org/citation.cfm?id=2512956

25. Z. Zhao and Q. Mei; *Questions about questions: an empirical analysis of information needs on Twitter*. In Proceedings of the 22nd international conference on World Wide Web (WWW), pp. 1545-1556, 2013. http://dl.acm.org/citation.cfm?id=2488523

26. X. Gao, E. Ferrara, J. Qiu. *Parallel Clustering of High-Dimensional Social Media Data Streams*. 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, 2015. http://arxiv.org/abs/1502.00316

27. S. Carton, S. Park, N. Zeffer, E. Adar, Q. Mei, and P. Resnick; *Audience Analysis for Competing Memes in Social Media.* ICWSM 2015. http://www.cond.org/rumorlens_icwsm_2015_final.pdf

28. V. S. Subrahmanian, O. Varol, P. Shiralkar, E. Ferrara, F. Menczer, A. Flammini, et al.; *The DARPA Twitter Bot Challenge*. IEEE Computer (to appear), 2015.

29. G. L. Ciampaglia, P. Shiralkar, L. M. Rocha, J. Bollen, F. Menczer, and A. Flammini; *Computational Fact Checking from Knowledge Networks*. PLoS ONE 10(6):e0128193, 2015 http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0128193

30. L. Weng, F. Menczer, and A. Flammini; *Online Interactions*. To appear in "Social Phenomena: From Data to Models" edited by B. Goncalves & N. Perra, Springer, Sep 2015 http://www.springer.com/us/book/9783319140100